GB97/03032

The Patent Office

The Patent Office
Cardiff Road
Newport
Gwent
NP9 1RH

5

09/297703

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation and Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

PRIORITY DOCUMENT

Signed

Dated 21 · 11 · 1997

Patents Form 1/77

Patents Act 1977
(Rule 16)

The Patent Office

06NOV96 E232121-1 002819
P01/7700 25.00

THE PATENT OFFICE
E
05 NOV 1996
NEWPORT

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

The Patent Office

Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference

   HCM/MJL/C244/N

2. Patent application number
   (The Patent Office will fill in this par

   9623095.8

   05 NOV 1996

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

   National Starch and Chemical
   Investment Holding Corporation
   Suite 27, 501 Silverside Road
   Wilmington, Delaware 19809
   United States of America

   Patents ADP number *(if you know it)*

   If the applicant is a corporate body, give the country/state of its incorporation

   Delaware, USA

   677169Sea1

4. Title of the invention

   Improvements in or Relating to Starch Content of Plants

5. Name of your agent *(if you have one)*

   Keith W Nash & Co

   "Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

   90-92 Regent Street
   Cambridge
   CB2 1DP

   OUT OF SEQUENCE
   DATE 05 NOV 1996

   Patents ADP number *(if you know it)*

   1206001

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day / month / year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*

   a) any applicant named in part 3 is not an inventor, or
   b) there is an inventor who is not named as an applicant, or
   c) any named applicant is a corporate body.
   See note (d))

   YES

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	
Description	19
Claim(s)	4
Abstract	1
Drawing(s)	16

10. If you are also filing any of the following, state how many against each item.

Priority documents	
Translations of priority documents	
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	
Request for preliminary examination and search *(Patents Form 9/77)*	
Request for substantive examination *(Patents Form 10/77)*	
Any other documents *(please specify)*	

11. I/We request the grant of a patent on the basis of this application.

Signature Keith W. Nash Date 4/11/96

Keith W Nash & Co

12. Name and daytime telephone number of person to contact in the United Kingdom

M J LIPSCOMBE (01223) 355477

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.

b) Write your answers in capital letters using black ink or you may type them.

c) If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.

d) If you have answered 'Yes' Patents Form 7/77 will need to be filed.

e) Once you have filled in the form you must remember to sign and date it.

f) For details of the fee and ways to pay please contact the Patent Office.

C244.00/N

Title: **Improvements in or Relating to Starch Content of Plants**

Field of the Invention

This invention relates to novel nucleic acid sequences, vectors and host cells comprising the nucleic acid sequence(s), to polypeptides encoded thereby, and to a method of altering a host cell by introducing the nucleic acid sequence(s) of the invention.

Background to the Invention

Starch consists of two main polysaccharides, amylose and amylopectin. Amylose is a linear polymer containing α-1,4 linked glucose units, while amylopectin is a highly branched polymer consisting of a α-1,4 linked glucan backbone with α-1,6 linked glucan branches. In most plant storage reserves amylopectin consitutes about 75% of the starch content. Amylopectin is synthesized by the concerted action of soluble starch synthase and starch branching enzyme [α-1,4 glucan: α-1,4 glucan 6-glycosyltransferase, EC 2.4.1.18]. Starch branching enzyme (SBE) hydrolyses α-1,4 linkages and rejoins the cleaved glucan, via an α-1,6 linkage, to an acceptor chain to produce a branched structure. The physical properties of starch are strongly affected by the relative abundance of amylose and amylopectin, and SBE is therefore a crucial enzyme in determining both the quantity and quality of starches produced in plant systems.

Starches are commercially available from several plant sources including maize, potato and cassava. Each of these starches has unique physical characteristics and properties and a variety of possible industrial uses. In maize there are a number of naturally occurring mutants which have altered starch composition such as high amylopectin types ("waxy" starches) or high amylose starches but in potato and cassava no such mutants exist on a commercial basis as yet.

Genetic modification offers the possibility of obtaining new starches which may have novel

and potentially useful characteristics. Most of the work to date has involved potato plants because they are amenable to genetic manipulation i.e. they can be transformed using Agrobacterium and regenerated easily from tissue culture. In addition many of the genes involved in starch biosynthesis have been cloned from potato and thus are available as targets for genetic manipulation, for example, by antisense inhibition of expression or sense suppression.

Cassava (*Manihot esculenta* L. Crantz) is an important crop in the tropics, where its starch-filled roots are used both as a food source and increasingly as a source of starch. Cassava is a high yielding perennial crop that can grow on poor soils and is also tolerant of drought. Cassava starch being a root-derived starch has properties similar but not identical to potato starch and is composed of 20-25% amylose and 75-80% amylopectin (Rickard *et al.*, 1991. Trop. Sci. **31**, 189-207). Some of the genes involved in starch biosynthesis have been cloned from cassava, including starch branching enzyme I (SBE I) (Salehuzzaman *et al.*, 1994 Plant Science **98**, 53-62), and granule bound starch synthase I (GBSS I) (Salehuzzaman *et al.*, 1993 Plant Molecular Biology **23**, 947-962) and some work has been done on their expression patterns although only in *in vitro* grown plants (Salehuzzaman *et al.*, 1994 Plant Science **98**, 53-62).

In most plants studied to date e.g. maize (Boyer & Preiss, 1978 Biochem. Biophys. Res. Comm. *80*, 169-175), rice (Smyth, 1988 Plant Sci. *57*, 1-8) and pea (Smith, Planta *175*, 270-279), two forms of SBE have been identified, each encoded by a separate gene. A recent review by Burton *et al.*, (1995 The Plant Journal *7*, 3-15) has demonstrated that the two forms of SBE constitute distinct classes of the enzyme such that, in general, enzymes of the same class from different plants may exhibit greater similarity than enzymes of different classes from the same plant. In their review, Burton *et al.* termed the two respective enzyme families class "A" and class "B", and the reader is referred thereto (and to the references cited therein) for a detailed discussion of the distinctions between the two classes. One general distinction of note would appear to be the presence, in class A SBE molecules, of a flexible N-terminal domain, which is not found in class B molecules. The distinctions noted by Burton *et al.* are relied on herein to define class A and class B SBE molecules, which terms are to be interpreted accordingly.

Many organisations have interests in obtaining modified Cassava starches by means of genetic modification. This is impossible to achieve however, unless the plant is amenable to transformation and regeneration, and the starch biosynthesis genes which are to be targeted for modification must be cloned. The production of transgenic cassava plants has only recently been demonstrated (Taylor *et al.*, 1996 Nature Biotechnology **14**, 726-730; Schöpke *et al.*, 1996 Nature Biotechnology **14**, 731-735; and Li *et al.*, 1996 Nature Biotechnology **14**, 736-740). The present invention concerns the identification, cloning and sequencing of a starch biosynthetic gene from Cassava, suitable as a target for genetic manipulation.

Summary of the Invention

In a first aspect the invention provides a nucleic acid sequence encoding a polypeptide having starch branching enzyme (SBE) activity, the polypeptide comprising an effective portion of amino acid residues 1-836 of the sequence shown in Figure 4. The nucleic acid is conveniently in substantial isolation, especially in isolation from other naturally associated nucleic acid sequences.

An "effective portion" of amino acid residues 1-836 may be defined as a portion which retains sufficient SBE activity when expressed in *E. coli* KV832 to complement the branching enzyme mutation therein. The amino acid sequence shown in Figure 4 includes the N terminal transit peptide, which comprises about the first 50 amino acid residues. As those skilled in the art will be well aware, such a transit peptide is not essential for SBE activity. Thus the mature polypeptide, lacking a transit peptide, may be considered as one example of an effective portion of residues 1-836.

Other effective portions may be obtained by effecting minor deletions in the amino acid sequence, whilst substantially preserving SBE activity. Comparison with known class A SBE sequences, with the benefit of the disclosure herein, will enable those skilled in the art to identify regions of the polypeptide which are less well conserved and so amenable to minor deletion, or amino acid substitution (particularly, conservative amino acid substitution) whilst substantially preserving SBE activity. Such less well-conserved

regions are generally found in the N terminal 179 amino acid residues (up to the triple proline "elbow" at residues 180-183) and in the last 50 residues or so of the C terminal, and in particular in the acidic tail of the C terminal.

Conveniently the nucleic acid sequence is obtainable from cassava, preferably obtained therefrom, and typically encodes a polypeptide obtainable from cassava. In a particular embodiment, the encoded polypeptide may have the amino acid sequence NSKH at about position 697, which sequence appears peculiar to an isoform of the SBE class A enzyme of cassava, other class A SBE enzymes having the conserved sequence DA D/E Y (Burton *et al.*, 1995 cited above).

In a particular embodiment the nucleic acid comprises a portion of nucleotides 21 to 2531 of the nucleic acid sequence shown in Figure 4, or a functionally equivalent nucleic acid sequence. Such functionally equivalent nucleic acid sequences include sequences which encode substantially the same polypeptide, but which differ in nucleotide sequence from that shown in Figure 4 by virtue of the degeneracy of the genetic code. For example, a nucleic acid sequence may be altered (e.g. "codon optimised") for expression in a host other than cassava, such that the nucleotide sequence differs substantially whilst the amino acid sequence of the encoded polypeptide is unchanged. Other functionally equivalent nucleic acid sequences are those which will hybridise under stringent hybridisation conditions (e.g. as described by Sambrook *et al.*, Molecular Cloning. A Laboratory Manual, CSH, i.e. washing with 0.1xSSC, 0.5% SDS at 68°C) with the sequence shown in Figure 4. Figure 10 shows a functionally equivalent sequence designated "125 + 94", which includes a region corresponding to the 3' coding portion of the sequence in Figure 4.

Functionally equivalent DNA sequences will preferably comprise at least 200-300bp, more preferably 300-600bp, and will exhibit at least 90% identity (preferably at least 95% identity) with the corresponding region of the DNA sequence shown in figures 4 or 10. Those skilled in the art will readily be able to conduct a sequence alignment between the putative functionally equivalent sequence and those detailed in Figures 4 or 10 - the identity of the two sequences is to be compared in those regions which are aligned by

standard computer software, which aligns corresponding regions of the sequences.

In particular embodiments the nucleic acid sequence may alternatively comprise a 5' and/or a 3' untranslated region ("UTR"), examples of which are shown in Figures 2 and 4. Figure 9 includes a 3' UTR, as nucleotides 688-1044 and Figure 10 includes 3' UTR as nucleotides 1507-1900 (which nucleotides correspond to the first base after the "stop" codon to the base immediately preceding the poly (A) tail). Any one of the sequences defined above, or a functional equivalent thereof (as defined by hybridisation properties, as set out in the preceding paragraph), could be useful in sense or anti-sense inhibition of corresponding genes, as will be apparent to those skilled in the art. It will also be apparent to those skilled in the art that such regions may be modified so as to optimise expression in a particular type of host cell and that the 5' and/or 3' UTRs could be used in isolation, or in combination with a coding portion of the sequence of the invention. Similarly, a coding portion could be used without a 5' or a 3' UTR if desired.

In a further aspect, the invention provides a replicable nucleic acid construct comprising any one of the nucleic acid sequences defined above. The construct will typically comprise a selectable marker and may allow for expression of the nucleic acid sequence of the invention. Conveniently the vector will comprise a promoter (especially a promoter sequence operable in a plant and/or a promoter operable in a bacterial cell) and one or more regulatory signals known to those skilled in the art.

In another aspect the invention provides a polypeptide having SBE activity, the polypeptide comprising an effective portion of amino acid residues 1-863 of the amino acid sequence shown in Figure 4. The polypeptide is conveniently one obtainable from cassava, although it may be derived using recombinant DNA techniques. The polypeptide is preferably in substantial isolation from other polypeptides, especially in isolation from polypeptides of plant origin. The polypeptide may have amino acid residues NSKH at about position 697, instead of the sequence DA D/E Y found in other SBE class A polypeptides. The polypeptide may be used in a method of modifying starch *in vitro*, the method comprising treating starch under suitable conditions (of temperature, pH etc.) with an effective amount of the polypeptide.

Those skilled in the art will appreciate that the disclosure of the present specification can be utilised in a number of ways. In particular, the characteristics of a host cell may be altered by recombinant DNA techniques. Thus, in a further aspect, there is provided a method by which a host cell may be altered by introduction of a nucleic acid sequence comprising at least 200bp and exhibiting at least 90% sequence identity with the corresponding region of the DNA sequence shown in Figures 4, 9 or 10, operably linked in the sense or (preferably) in the anti-sense orientation to a suitable promoter active in the host cell, and causing transcription of the introduced nucleic acid sequence, said transcript and/or the translation product thereof being sufficient to interfere with the expression of a homologous gene naturally present in said host cell, which homologous gene encodes a polypeptide having SBE activity. The altered host cell is typically a plant cell, such as a cell of a cassava, banana, potato, sweet potato, tomato, pea, wheat, barley, oat, maize, or rice plant.

Desirably the method further comprises the introduction of one or more nucleic acid sequences which are effective in interfering with the expression of other homologous gene or genes naturally present in the host cell. Such other genes whose expression is inhibited may be involved in starch biosynthesis (e.g. an SBE I gene), or may be unrelated to SBE II.

Those skilled in the art will be aware that both anti-sense inhibition, and "sense suppression" of expression of genes, especially plant genes, has been demonstrated (e.g. Matzke & Matzke 1995 Plant Physiol. 107, 679-685).

It is believed that antisense methods are mainly operable by the production of antisense mRNA which hybridises to the sense mRNA, preventing its translation into functional polypeptide, possibly by causing the hybrid RNA to be degraded (e.g. Sheehy *et al.*, 1988 PNAS *85*, 8805-8809; Van der Krol *et al.*, Mol. Gen. Genet. *220*, 204-212). Sense suppression also requires homology between the introduced sequence and the target gene, but the exact mechanism is unclear. It is apparent however that, in relation to both antisense and sense suppression, neither a full length nucleotide sequence, nor a "native" sequence is essential. Preferably the nucleic acid sequence used in the method will

comprise at least 200-300bp, more preferably at least 300-600bp, of the full length sequence, but by simple trial and error other fragments (smaller or larger) may be found which are functional in altering the characteristics of the plant. It is also known that untranslated portions of sequence can suffice to inhibit expression of the homologous gene - coding portions may be present within the introduced sequence, but they do not appear to be essential under all circumstances.

The inventors have discovered that there are at least two class A SBE genes in cassava. A fragment of a second gene has been isolated, which fragment directs the expression of the C terminal 481 amino acids of cassava class A SBE (see Figure 10) and comprises a 3' untranslated region. The coding portions of the two genes show some differences, and that portion of the fragmentary SBE gene may be considered as functionally equivalent to the corresponding portion of the nucleotide sequence shown in Figure 4. However, the 3' untranslated regions of the two genes show marked differences. Thus the method of altering a host cell may comprise the use of a sufficient portion of either gene so as to inhibit the expression of the naturally occurring homologous gene. Conveniently, a portion of nucleotide sequence is employed which is conserved between both genes. Alternatively, sufficient portions of both genes may be employed, typically using a single construct to direct the transcription of both introduced sequences.

In addition, as explained above, it may be desired to cause inhibition of expression of the class B SBE (i.e. SBE I) in the same host cell. A number of class B SBE gene sequences are known, including portions of the cassava class B SBE (Salehuzzaman *et al.*, 1994 Plant Science *98*, 53-62) and any one of these may prove suitable. Preferably the sequence used is that which derives from the host cell sought to be altered (e.g. when altering the characteristics of a cassava plant cell, it is generally preferred to use sense or anti-sense sequences corresponding exactly to at least portions of the cassava gene whose expression is sought to be inhibited).

In a further aspect the invention provides an altered host cell, into which has been introduced a nucleic acid sequence comprising at least 200bp and exhibiting at least 90% sequence identity with the corresponding region of the DNA sequence shown in Figures

4, 9 or 10, operably linked in the sense or anti-sense orientation to a suitable promoter, said host cell comprising a natural gene sharing sequence homology with the introduced sequence.

The host cell may be a micro-organism (such as a bacterial, fungal or yeast cell) or a plant cell. Conveniently the host cell altered by the method is a cell of a cassava plant, or another plant with starch storage reserves, such as banana, potato, sweet potato, tomato, pea, wheat, barley, oat, maize, or rice plant. Typically the sequence will be introduced in a nucleic acid construct, by way of transformation, transduction, micro-injection or other method known to those skilled in the art. The invention also provides for a plant into which has been introduced a nucleic acid sequence of the invention, or the progeny of such a plant.

The altered plant cell will preferably be grown into an altered plant, using techniques of plant growth and cultivation well-known to those skilled in the art of re-generating plantlets from plant cells.

The invention also provides a method of obtaining starch from an altered plant, the plant being obtained by the method defined above. Starch may be extracted from the plant by any of the known techniques (e.g. milling). The invention further provides starch obtainable from a plant altered by the method defined above, the starch having altered properties compared to starch extracted from an equivalent but unaltered plant. Conveniently the altered starch is obtained from an altered plant selected from the group consisting of cassava, potato, pea, tomato, maize, wheat, barley, oat, sweet potato and rice. Typically the altered starch will have increased amylose content.

The invention will now be further described by way of illustrative examples and with reference to the accompanying drawings, in which:-

Figure 1 is a schematic illustration of the cloning strategy for cassava SBE II. The top line represents the size of a full length clone with distances in kilobases (kb) and arrows representing oligonucleotides (rightward pointing arrows are sense strand, leftward are on

opposite strand). The long thick arrow is the open reading frame with start and stop codons shown. Below this are shown the 3' RACE, 5' RACE and PCR clones identified either by the plasmid name (shown in brackets above the line) or the clone number (shown to the left of the clone) for the 5' RACE only. Also shown (by an x) in the 5' RACE clones are positions of small deletions or introns.

Figure 2 shows the DNA sequence and predicted ORF of csbe2con.seq. This sequence is a consensus of 3' RACE pSJ94 and 5' RACE clones 27/9,11 and 28. The first 64 base pairs are derived from the RoRidT17 adaptor primer/dT tail followed by the SBE sequence. The one long open reading frame is shown in one letter code below the double strand DNA sequence. Also shown is the upstream ORF (MQL...LPW).

Figure 3 shows an alignment of the 5' region of cassava SBE II csbe2con and pSJ99 (clones 20 and 35) DNA sequences. Differences from the consensus sequence are shaded.

Figure 4 shows the DNA sequence and predicted ORF of full length cassava SBE II tuber cDNA in pSJ107. The sequence shown is from the CSBE214 to the CSBE218 oligonucleotide.

Figure 5 shows an alignment of 3' region of cassava SBE II pSJ116 and 125+94 DNA sequences. The top line is the 125 + 94 sequence and the bottom SJ116 sequence. Identical nucleotides are indicated by the same letter in the middle line, differences are indicated by a gap, and dashed lines indicate gaps introduced to optimise alignment.

Figure 6 shows an alignment of carboxy terminal region of pSJ116 and 125+94 protein sequences. The top sequence is from 125+94 and the bottom from pSJ116. Identical amino acid residues are shown with the same letter, conserved changes with a colon and neutral changes with a period.

Figure 7 shows a phylogenetic tree of starch branching enzyme proteins. The length of each pair of branches represents the distance between sequence pairs. The scale beneath the tree measures the distance between sequences (units indicate the number of substitution

events). Dotted lines indicate a negative branch length because of averaging the tree. Zmcon12.pro is maize SBE II, psstb1.pro is pea SBE I (Bhattacharyya *et al* 1990 Cell **60**, 115-121) and atsbe2-1 & 2-2.pro are two SBE II proteins from *Arabidopsis thalania* (Fisher *et al* 1996 Plant Mol. Biol. **30**, 97-108). SJ107.pro is representative of a cassava SBE II sequence, and potsbe2.pro is a potato SBE II sequence known to the inventors.

Figure 8 is an alignment of SBE II proteins. Protein sequences are indicated in one letter code. The top line represents the consensus sequence, below which is shown the consensus ruler and the individual SBE II sequences. Residues matching the consensus are shaded. Dashes represent gaps introduced to optimise alignment. Sequence identities are shown at the right of the figure and are as Figure 7, except that SJ107.pro is cassava SBE II.

Figure 9 shows the DNA sequence and predicted ORF of a cassava SBE II cDNA isolated by 3' RACE (plasmid pSJ 101).

Figure 10 shows the consensus DNA sequence and predicted ORF of a second cassava SBE II cDNA isolated by 3' and 5' RACE (sequence designated 125+94 is from plasmid pSJ125 and pSJ94, spliced at the CSBE217 oligo sequence).

Figure 11 is a schematic diagram of the plant transformation vector pSJ64. The black line represents the DNA sequence. The hashed line represents the bacterial plasmid backbone (containing the origin of replication and bacterial selection marker) and is not shown in full. The filled triangles represent the T-DNA borders (RB = right border, LB = left border). Relevant restriction enzyme sites are shown above the black line with the approximate distances (in kiloobases) betwen sites marked by an asterisk shown underneath. The thinnest arrows represent polyadenylation signals (pAnos = nopaline synthase, pAg7 = Agrobacterium gene 7), the intermediate arrows represent protein coding regions (SBE II = cassava SBE II, HYG = hygromycin resistance gene) and the thick arrows represent promoter regions (P-2x35S = double CaMV 35S promoter, P-nos = nopaline synthase promoter).

Example 1

This example relates to the isolation and cloning of SBE II sequences from cassava.

Recombinant DNA manipulations

Standard procedures were performed essentially according to Sambrook *et al.* (1989 Molecular cloning A laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.). DNA sequencing was performed on an ABI automated DNA sequencer and sequences manipulated using DNASTAR software for the Macintosh.

Rapid Amplification of cDNA ends (RACE) and PCR conditions

5' and 3' RACE were performed essentially according to Frohman *et al.*, (1988 Proc. Natl. Acad. Sci. USA **85**, 8998-9002) but with the following modifications.

For 3' RACE, 5 μg of total RNA was reverse transcribed using 5 pmol of the RACE adaptor RoRidT17 as primer and Stratascript RNAse H- reverse transcriptase (50 U) in a 50 μl reaction according to the manufacturer's instructions (Stratagene). The reaction was incubated for 1 hour at 37°C and then diluted to 200 μl with TE (10 mM Tris HCl, 1 mM EDTA) pH 8 and stored at 4°C. 2.5 μl of this cDNA was used in a 25 μl PCR reaction with 12.5 pmol of SBE A and Ro primers for 30 cycles of 94°C 45 sec, 50°C 25 sec, 72°C 1 min 30 sec. A second round of PCR (25 cycles) was performed using 1 μl of this reaction as template in a 50 μl reaction under the same conditions. Amplified products were separated by agarose gel electrophoresis and cloned into the pT7Blue vector (Invitrogen).

For the first round of 5' RACE, 5 μg of total leaf RNA was reverse transcribed as described above using 10 pmol of the SBE II gene specific primer CSBE22. This primer was removed from the reaction by diluting to 500 μl with TE and centrifuging twice through a centricon 100 microconcentrator. The concentrated cDNA was then dA-tailed with 9U of terminal deoxynucleotide transferase and 50 μM dATP in a 20 μl reaction in buffer supplied by the manufacturer (BRL). The reaction was incubated for 10 min at 37°C and 5 min at 65°C and then diluted to 200 μl with TE pH 8. PCR was performed in a 50 μl volume using 5μl of tailed cDNA, 2.5 pmol of RoRidT17 and 25 pmol of Ro

and CSBE24 primers for 30 cycles of 94°C 45 sec, 55°C 25 sec, 72°C 3 min. Amplified products were separated on a 1% TAE agarose gel, cut out, 200μl of TE was added and melted at 99°C for 10 min. Five μl of this was re-amplified in a 50 μl volume using CSBE25 and Ri as primers and 25 cycles of 94°C 45 sec, 55°C 25 sec, 72°C 1 min 30 sec. Amplified fragments were separated on a 1% TAE agarose gel, purified on DEAE paper and cloned into pT7Blue.

The second round of 5' RACE was performed using CSBE28 and 29 primers in the first and second round PCR reactions respectively using a new A-tailed cDNA library primed with CSBE27.

A third round of 5' RACE was performed on the same CSBE27 primed cDNA .

Repeat 3' RACE and PCR Cloning

The 3' RACE library (RoRidT17 primed leaf RNA) was used as a template. The first PCR reaction was diluted 1:20 and 1 μl was used in a 50 μl PCR reaction with SBE A and Ri primers and the products were cloned into pT7Blue. The cloned PCR products were screened for the presence or absence of the CSBE23 oligo by colony PCR.

A full length cDNA of cassava SBE II was isolated by PCR from leaf or root cDNA (RoRidT17 primed) using primers CSBE214 and CSBE218 from 2.5 μl of cDNA in a 25 μl reaction and 30 cycles of 94°C 45 sec, 55°C 25 sec, 72°C 2 min.

Complementation of *E. coli* mutant KV832

SBE II containing plasmids were transformed into the branching enzyme deficient mutant *E. coli* KV832 (Keil *et al.*, 1987 Mol. Gen. Genet. **207**, 294-301) and cells grown on solid PYG media (0.85 % KH_2PO_4, 1.1 % K_2HPO_4, 0.6 % yeast extract) containing 1.0 % glucose. To test for complementation, a loop of cells was scraped off and resuspended in 150 μL water to which was added 15 μL of Lugol's solution (2 g KI and 1 g I_2 per 300 ml water).

RNA isolation

RNA was isolated from cassava plants by the method of Logemann (1987 Anal. Biochem. **163**, 21-26). Leaf RNA was isolated from 0.5 gm of in vitro grown plant tissue. The total yield was 300 μg. Three month old roots (88 gm) were used for isolation of root RNA).

SBE II specific oligonucleotides

SBE A	ATGGACAAGGATATGTATGA
CSBE21	GGTTTCATGACTTCTGAGCA
CSBE22	TGCTCAGAAGTCATGAAACC
CSBE23	TCCAGTCTCAATATACGTCG
CSBE24	AGGAGTAGATGGTCTGTCGA
CSBE25	TCATACATATCCTTGTCCAT
CSBE26	GGGTGACTTCAATGATGTAC
CSBE27	GGTGTACATCATTGAAGTCA
CSBE28	AATTACTGGCTCCGTACTAC
CSBE29	CATTCCAACGTGCGACTCAT
CSBE210	TACCGGTAATCTAGGTGTTG
CSBE211	GGACCTTGGTTTAGATCCAA
CSBE212	ATGAGTCGCACGTTGGAATG
CSBE213	CAACACCTAGATTACCGGTA
CSBE214	TTAGTTGCGTCAGTTCTCAC
CSBE215	AATATCTATCTCAGCCGGAG
CSBE216	ATCTTAGATAGTCTGCATCA
CSBE217	TGGTTGTTCCCTGGAATTAC
CSBE218	TGCAAGGACCGTGACATCAA

RESULTS

Cloning of a SBE II gene from cassava leaf

The strategy for cloning a full length cDNA of starch branching enzyme II of cassava is shown in Figure 1. A comparison of several SBE II (class A) SBE DNA sequences identified a 23 bp region which appears to be completely conserved among most genes (data not shown) and is positioned about one kilobase upstream from the 3' end of the

gene. An oligonucleotide primer (designated SBE A) was made to this sequence and used to isolate a partial cDNA clone by 3' RACE PCR from first strand leaf cDNA as illustrated in Figure 1. An approximately 1100 bp band was amplified, cloned into pT7Blue vector and sequenced. This clone was designated pSJ94 and contained a 1120 bp insert starting with the SBE A oligo and ending with a polyA tail. There was a predicted open reading frame of 235 amino acids which was highly homologous (79% identical) to a potato SBE II also isolated by the inventors (data not shown) suggesting that this clone represented a class A (SBE II) gene.

To obtain the sequence of a full length clone nested primers were made complementary to the 5' end of this sequence and used in 5' RACE PCR to isolate clones from the 5' region of the gene. A total of three rounds of 5' RACE was needed to determine the sequence of the complete gene (i.e. one that has a predicted long ORF preceded by stop codons). It should be noted that during this cloning process several clones (# 23, 9, 16) were obtained that had small deletions and in one case (clone 23) there was also a small (120 bp) intron present. These occurrences are not uncommon and probably arise through errors in the PCR process and/or reverse transcription of incompletely processed RNA (heterogeneous nuclear RNA).

The overlapping cDNA fragments could be assembled into a contiguous 3 kb sequence (designated csbe2con.seq) which contained one long predicted ORF as shown in Figure 2. Several clones in the last round of 5' RACE were obtained which included sequence of the untranslated leader (UTL). All of these clones had an ORF (42 amino acids) 46 bp upstream and out of frame with that of the long ORF.

There is more than one SBE II gene in cassava

In order to determine if the assembled sequence represented that of a single gene, attempts were made to recover by PCR a full length SBE II gene using primers CSBE214 and CSBE23 at the 5' and 3' ends of the csbe2con sequence respectively. All attempts were unsuccessful using either leaf or root cDNA as template. The PCR was therefore repeated with either the 5'- or 3'- most primer and complementary primers along the length of the SBE II gene to determine the size of the largest fragment that could be amplified. With

the CSBE214 primer, fragments could be amplified using primers 210, 28, 27 and 22 in order of increasing distance, the latter primer pair amplifying a 2.2 kb band. With the 3' primer CSBE23, only primer pairs with 21 and 26 gave amplification products, the latter being about 1200 bp. These results suggest that the original 3' RACE clone (pSJ94) is derived from a different SBE II gene than the rest of the 5' RACE clones even though the two largest PCR fragments (214+22 and 26+23) overlap by 750 bp and share several primer sites. It is likely that the sequence of the two genes starts to diverge around the CSBE22 primer site such that the 3' end of the corresponding gene does not contain the 23 primer and is not therefore able to amplify a cDNA when used with the 214 primer.

To confirm this, the sequence of the longest 5' PCR fragment (214+22) from two clones (#20 designated pSJ99, & #35) was determined and compared to the consensus sequence csbe2con as shown in Figure 3. The first 2000 bases are nearly identical (the single base changes might well be PCR errors), however the consensus sequence is significantly different after this. This region corresponds to the original 3' RACE fragment pSJ94 (SBE A + Ri adaptor) and provided evidence that there may be more than one SBE II gene in cassava.

The 3' end corresponding to pSJ99 was therefore cloned as follows: 3' RACE PCR was performed on leaf cDNA using the SBE A oligo as the gene specific primer so that all SBE II genes would be amplified. The cloned DNA fragments were then screened for the presence or absence of the CSBE23 primer by PCR. Two out of 15 clones were positive with the SBE A + Ri primer pair but negative with SBE A + CSBE23 primers. The sequence of these two clones (designated pSJ101, as shown in Figure 9) demonstrated that they were indeed from an SBE II gene and that they were different from pSJ94. However the overlapping region of pSJ101 (the 3' clone) and pSJ99 (the 5' clone) was identical suggesting that they were derived from the same gene.

To confirm this a primer (CSBE218) was made to a region in the 3' UTR (untranslated region) of pSJ101 and used in combination with CSBE214 primer to recover by PCR a full length cDNA from both leaf and root cDNA. These clones were sequenced and designated pSJ106 & pSJ107 respectively. The sequence and predicted ORF of pSJ107

is shown in Figure 4. The long ORF in plasmid pSJ106 was found to be interrupted by a stop codon (presumably introduced in the PCR process) approximately 1 kb from the 3' end of the gene, therefore another cDNA clone (designated pSJ116) was amplified in a separate reaction, cloned and sequenced. This clone had an intact ORF (data not shown). There were only a few differences in these two sequences (in the transit peptide aa 27- 41: YRRTSSCLSFNFKEA to DRRTSSCLSFIFKKAA and L831 in pSJ107 to V in pSJ116 respectively).

An additional 740bp of sequence of the gene corresponding to the pSJ94 clone was isolated by 5' RACE using the primers CSBE216 and 217, and was designated pSJ125. This sequence was combined with that of pSJ94 to form a consensus sequence "125 + 94", as shown in Figure 10. The sequence of this second gene is about 90% identical at the DNA and protein level to pSJ116, as shown in Figure 5 and 6, and is clearly a second form of SBE II in cassava. The 3' untranslated regions of the two genes are not related (data not shown).

It was also determined that the full length cassava SBE II genes (from both leaf and tuber) actually encode for active starch branching enzymes since the cloned genes were able to complement the glycogen branching enzyme deficient *E. coli* mutant KV832.

Main Findings

1) A full length cDNA clone of a starch branching enzyme II (SBE II) gene has been cloned from leaves and starch storing roots of cassava. This cDNA encodes a 836 amino acid protein (Mr 95 Kd) and is 86 % identical to pea SBE I over the central conserved domain.

2) There is more than one SBE II gene in cassava as a second partial SBE II cDNA was isolated which differs slightly in the protein coding region from the first gene and has no homology in the 3' untranslated region.

3) The isolated full length cDNA from both leaves and roots encodes an active SBE as it complements an *E. coli* mutant deficient in glycogen branching enzyme as assayed by

iodine staining.

We have shown that there are SBE II (Class A) gene sequences present in the cassava genome by isolating cDNA fragments using 3' and 5' RACE. From these cDNA fragments a consensus sequence of over 3 kb could be compiled which contained one long open reading frame (Figure 2) which is highly homologous to other SBE II (class A) genes (data not shown). It is likely that the consensus sequence does not represent that of a single gene since attempts to PCR a full length gene using primers at the 5' and 3' ends of this sequence were not successful. In fact screening of a number of leaf derived 3' RACE cDNAs showed that a second SBE II gene (clone designated pSJ101) was also expressed which is highly homologous within the coding region to the originally isolated cDNA (pSJ94) but has a different 3' UTR. A full length SBE II gene was isolated from leaves and roots by PCR using a new primer to the 3' end of this sequence and the original sequence at the 5' end of the consensus sequence. If the frequency of clones isolated by 3' RACE PCR reflects the abundance of the mRNA levels then this full length gene may be expressed at lower levels in the leaf than the pSJ94 clone (2 out of 15 were the former class, 13/15 the latter). It should be noted that each class is expressed in both leaves and roots as judged by PCR (data not shown). Sequence analysis of the predicted ORF of the leaf and root genes showed only a few differences (4 amino acid changes and one deletion) which could have arisen through PCR errors or, alternatively, there may be more than one nearly identical gene expressed in these tissues.

A comparison of all known SBE II protein sequences shows that the cassava SBE II gene is most closely related to the pea gene (Figure 8). The two proteins are 86.3% identical over a 686 amino acid range which extends from the triple proline "elbow" (Burton *et al.*, 1995 Plant J. 7, 3-15) to the conserved VVYA sequence immediately preceding the C-terminal extensions (data not shown). All SBE II proteins are conserved over this range in that they are at least 80% similar to each other. Remarkably however, the sequence conservation between the pea, potato and cassava SBE II proteins also extends to the N-terminal transit peptide, especially the first 12 amino acids of the precursor protein and the region surrounding the mature terminus of the pea protein (AKFSRDS). Because the proteins are so similar around this region it can be predicted that the mature terminus of

the cassava SBE II protein is likely to be GKSSHES. The precursor has a predicted molecular mass of 96 kD and the mature protein a predicted molecule mass of 91.3 kD. The cassava SBE II has a short acidic tail at the C-terminal although this is not as long or as acidic as that found in the pea or potato proteins. The significance of this acidic tail, if any, remains to be determined. One notable difference between the amino acid sequence of cassava SBE II and all other SBE II proteins is the presence of the sequence NSKH at around position 697 instead of the conserved sequence DAD/EY. Although this conserved region forms part of a predicted α-helix (number 8) of the catalytic $(\beta/\alpha)_8$ barrel domain (Burton et al 1995 cited previously), this difference does not abolish the SBE activity of the cassava protein as this gene can still complement the glycogen branching deletion mutant of *E. coli*. It may however affect the specificity of the protein. An interesting point is that the other cassava SBE II clone pSJ94 has the conserved sequence DADY.

One other point of interest concerning the sequence of the SBE II gene is the presence of an upstream ATG in the 5' UTL. This ATG could initiate a small peptide of 42 amino acids which would terminate downstream of the predicted initiating methionine codon of the SBE II precursor. If this does occur then the translation of the SBE II protein from this mRNA is likely to be inefficient as ribosomes normally initiate at the 5' most ATG in the mRNA. However the first ATG is in a poorer Kozak context than the SBE II initiator and it may be too close to the 5' end of the message to initiate efficiently (14 nucleotides) thus allowing initiation to occur at the correct ATG.

In conclusion we have shown that cassava does have SBE II gene sequences, that they are expressed in both leaves and tubers and that more than one gene exists.

Example 2

Construction of plant transformation vectors and transformation of cassava with antisense starch branching enzyme genes.

This example describes in detail how a portion of the SBE II gene isolated from cassava may be introduced into cassava plants to create transgenic plants with altered properties.

An 1100 bp Hind III - Sac I fragment of cassava SBE II (from plasmid pSJ94) was cloned into the Hind III - Sac I sites of the plant transformation vector pSJ64 (Fig 11). This placed the SBE II gene in an antisense orientation between the 2X 35S CaMV promoter and the nopaline synthase polyadenylation signal. pSJ64 is a derivative of the binary vector pGPTV-HYG (Becker *et al.*, 1992 Plant Molecular Biology *20*: 1195-1197) modified by inclusion of an approximately 750 bp fragment of pJIT60 (Guerineau *et al* 1992 Plant Mol. Biol. *18*, 815-818) containing the duplicated cauliflower mosaic virus (CaMV) 35S promoter (Cabb-JI strain, equivalent to nucleotides 7040 to 7376 duplicated upstream of 7040 to 7433, as described by Frank *et al.*, 1980 Cell 21, 285-294) to replace the GUS coding sequence. A similar construct was made with the cassava SBE II sequence from plasmid pSJ101.

These plasmids are then introduced into *Agrobacterium tumefaciens* LBA4404 by a direct DNA uptake method (An *et al*, Binary vectors, In: Plant Molecular Biology Manual (ed Galvin and Schilperoort) AD 1988 pp 1-19) and can be used to transform cassava somatic embryos by selecting on hygromycin as described by Li *et al.* (1996, Nature Biotechnology *14*, 736-740).

Claims

1. A nucleic acid sequence encoding a polypeptide having starch branching enzyme (SBE) activity, the encoded polypeptide comprising at least an effective portion of amino acid residues 1-836 of the sequence shown in Figure 4.

2. A nucleic acid sequence according to claim 1, encoding a polypeptide comprising amino acid residues 1-836 of the sequence shown in Figure 4.

3. A nucleic acid sequence according to claim 1, comprising nucleotides 21-2531 of the nucleic acid sequence shown in Figure 4, or a functionally equivalent nucleotide sequence which hybridises under stringent hybridisation conditions with the nucleic acid sequence shown in Figure 4.

4. A nucleic acid sequence according to any one of claims 1, 2 or 3 comprising a 5' and/or a 3' untranslated region.

5. A nucleic acid sequence according to any one of the preceding claims, encoding a polypeptide having the amino acid sequence NSKH at about residue 697.

6. A nucleic acid sequence comprising at least 200bp and exhibiting at least 90% sequence identity with the corresponding region of the DNA sequence shown in Figures 4, 9 or 10, operably linked in the sense or anti-sense orientation to a promoter operable in plants.

7. A nucleic acid sequence according to claim 6, comprising at least 300-600bp.

8. A sequence according to claim 6 or 7, comprising a 5'and/or 3'untranslated region.

9. A sequence according to claim 8, comprising nucleotides 688-1044 of the sequence shown in Figure 9, and/or nucleotides 1507-1900 of the sequence shown in Figure 10.

10. A sequence according to claim 6, comprising the nucleotide sequence shown in Figure 10.

11. A replicable nucleic acid construct comprising a nucleic acid sequence according to any one of the preceding claims.

12. A polypeptide having SBE activity and comprising an effective portion of amino acid residues 1-863 of the amino acid sequence shown in Figure 4.

13. A polypeptide according to claim 12, in substantial isolation from other polypeptides.

14. A polypeptide according to claim 12 or 13, having the amino acid sequence NSKH at about position 697.

15. A method of modifying starch *in vitro*, the method comprising treating starch to be modified under suitable conditions with an effective amount of a polypeptide according to any one of claims 12, 13 or 14.

16. A method of altering a plant host cell, the method comprising introducing into the cell a nucleic acid sequence comprising at least 200bp and exhibiting at least 90% sequence identity with the corresponding region of the DNA sequence shown in Figures 4, 9 or 10, operably linked in the sense or anti-sense orientation to a suitable promoter active in the host cell, and causing transcription of the introduced nucleotide sequence, said transcript and/or the translation product thereof being sufficient to interfere with the expression of a homologous gene naturally present in the host cell, which homologous gene encodes a polypeptide having SBE activity.

17. A method according to claim 16, wherein the host cell is from a cassava, banana, potato, pea, tomato, maize, wheat, barley, oat, sweet potato or rice plant.

18. A method according to claim 16 or 17, comprising the introduction of one or more further nucleic acid sequences, operably linked in the sense or anti-sense orientation to a

suitable promoter active in the host cell, and causing transcription of the one or more further nucleic acid sequences, said transcripts and/or translation products thereof being sufficient to interfere with the expression of homologous gene(s) present in the host cell.

19. A method according to claim 18, wherein the one or more further nucleic acid sequences interfere with the expression of a gene involved in starch biosynthesis.

20. A method according to claim 18 or 19, wherein the further nucleic acid sequence comprises at least part of an SBE I gene.

21. A method according to claim 20, wherein the further nucleic acid sequence comprises at least part of the cassava SBE I gene.

22. A method according to any one of claims 16 - 21, wherein the host cell is selected from one of the following: cassava, banana, potato, pea, tomato, maize, wheat, barley, oat, sweet potato or rice.

23. A method according to any one of claims 16-22, wherein the altered host cell gives rise to starch having different properties compared to starch from an unaltered cell.

24. A method according to any one of claims 16-23, further comprising the step of growing the altered host cell into a plant or plantlet.

25. A method of obtaining starch having altered properties, comprising growing a plant from an altered host cell according to the method of claim 24, and extracting the starch therefrom.

26. A plant or plant cell into which has been artificially introduced a nucleic acid sequence comprising at least 200bp and exhibiting at least 90% sequence identity with the corresponding region of the DNA sequence shown in Figures 4, 9 or 10, operably linked in the sense or anti-sense orientation to a promoter operable in plants, or the progeny thereof.

27. A plant according to claim 24, altered by the method of any one of claims 16-22.

28. Starch obtainable from an altered plant according to claim 26 or 27, having altered properties compared to starch extracted from an equivalent but unaltered plant.

29. Starch obtained from an altered plant according to claim 26 or 27, having altered properties compared to starch extracted from an equivalent but unaltered plant.

30. Starch according to claim 28 or 29 obtained from an altered plant selected from the group consisting of:- cassava, banana, potato, pea, tomato, maize, wheat, barley, oat, sweet potato and rice plants.

31. Starch according to any one of claims 28, 29 or 30, having increased amylose content compared to starch extracted from an equivalent but unaltered plant.

ABSTRACT

Title: **Improvements in or Relating to Starch Content of Plants**

Disclosed is a nucleic acid sequence encoding a polypeptide having starch branching enzyme (SBE) activity, the encoded polypeptide comprising an effective portion of amino acid residues 1-836 of the sequence shown in Figure 4.

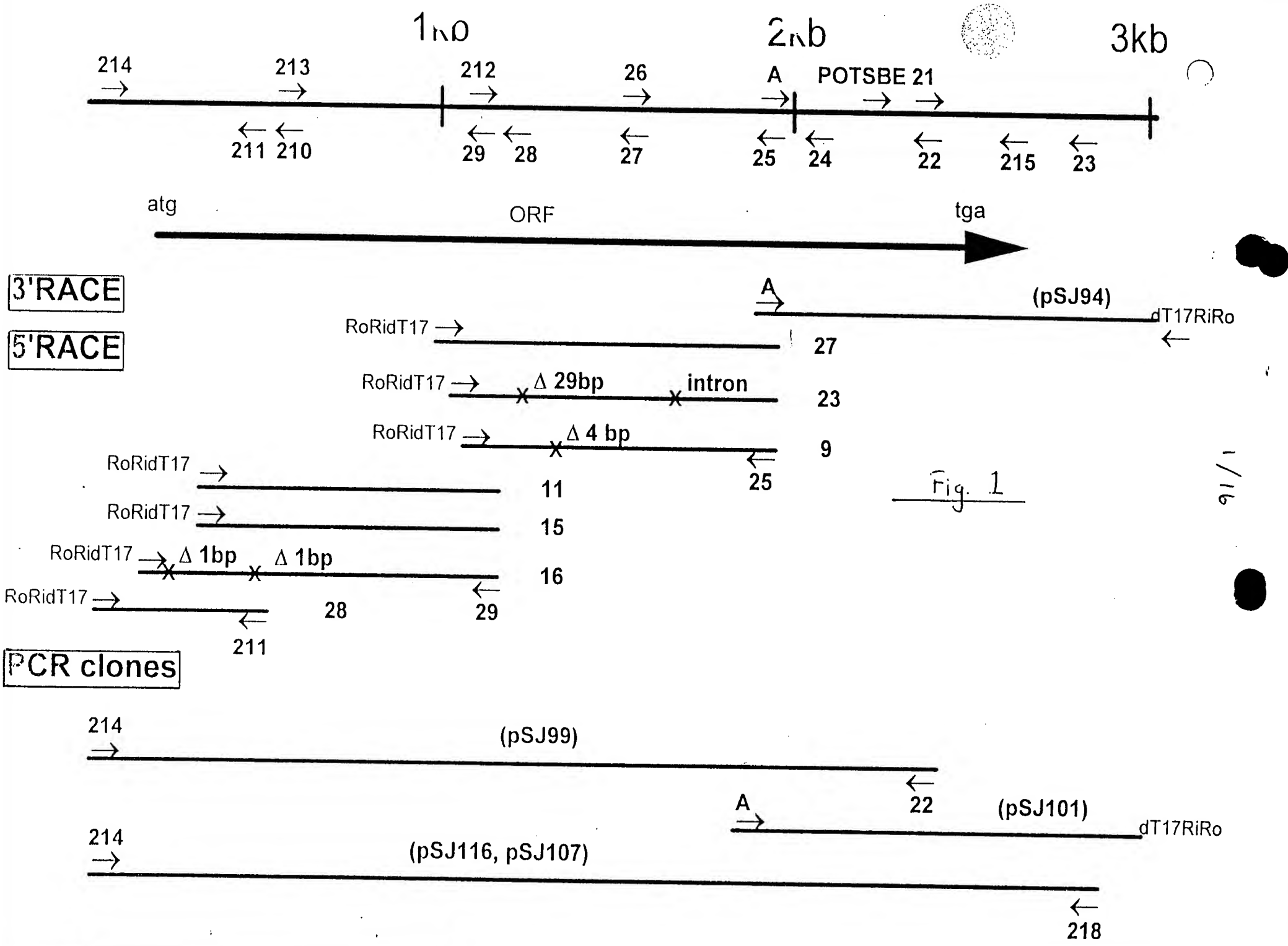

1kb
2kb
3kb
214
213
212
26
A
POTSBE 21
211 210
29 28
27
25 24
22
215
23
atg
ORF
tga
3'RACE
5'RACE
A
(pSJ94)
dT17RiRo
RoRidT17
27
RoRidT17
Δ 29bp
intron
23
RoRidT17
Δ 4 bp
9
25
RoRidT17
11
RoRidT17
15
RoRidT17
Δ 1bp
Δ 1bp
16
RoRidT17
28
29
211
PCR clones
214
(pSJ99)
22
A
(pSJ101)
dT17RiRo
214
(pSJ116, pSJ107)
218

Fig. 1

91/1

Fig. 2

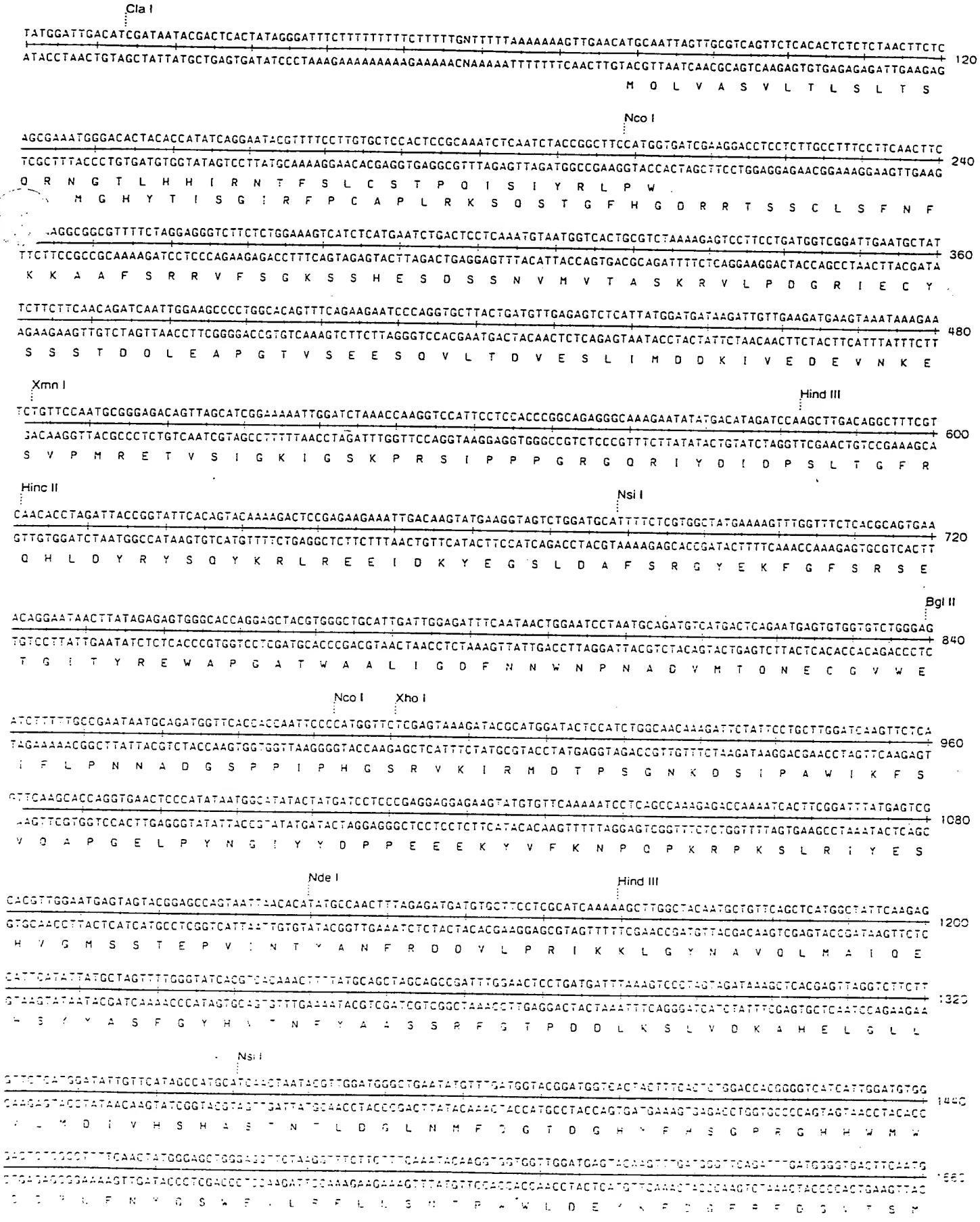

Cla I
Nco I
Xmn I
Hind III
Hinc II
Nsi I
Bgl II
Nco I
Xho I
Nde I
Hind III
Nsi I

Fig. 2 (con.)

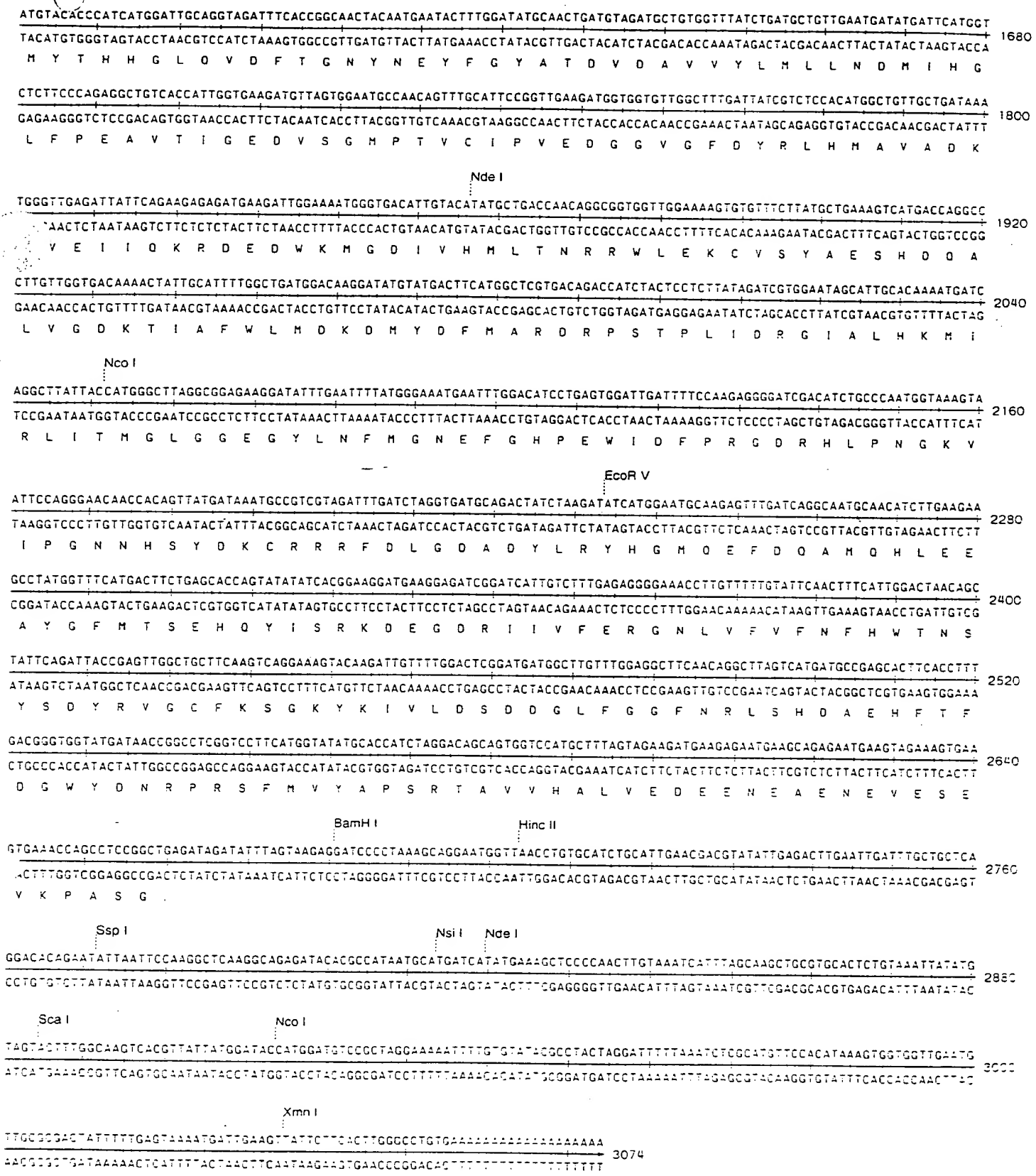

Nde I
Nco I
EcoR V
BamH I
Hinc II
Ssp I
Nsi I
Nde I
Sca I
Nco I
Xmn I
1680
1800
1920
2040
2160
2280
2400
2520
2640
2760
2880
3000
3074

Fig. 3

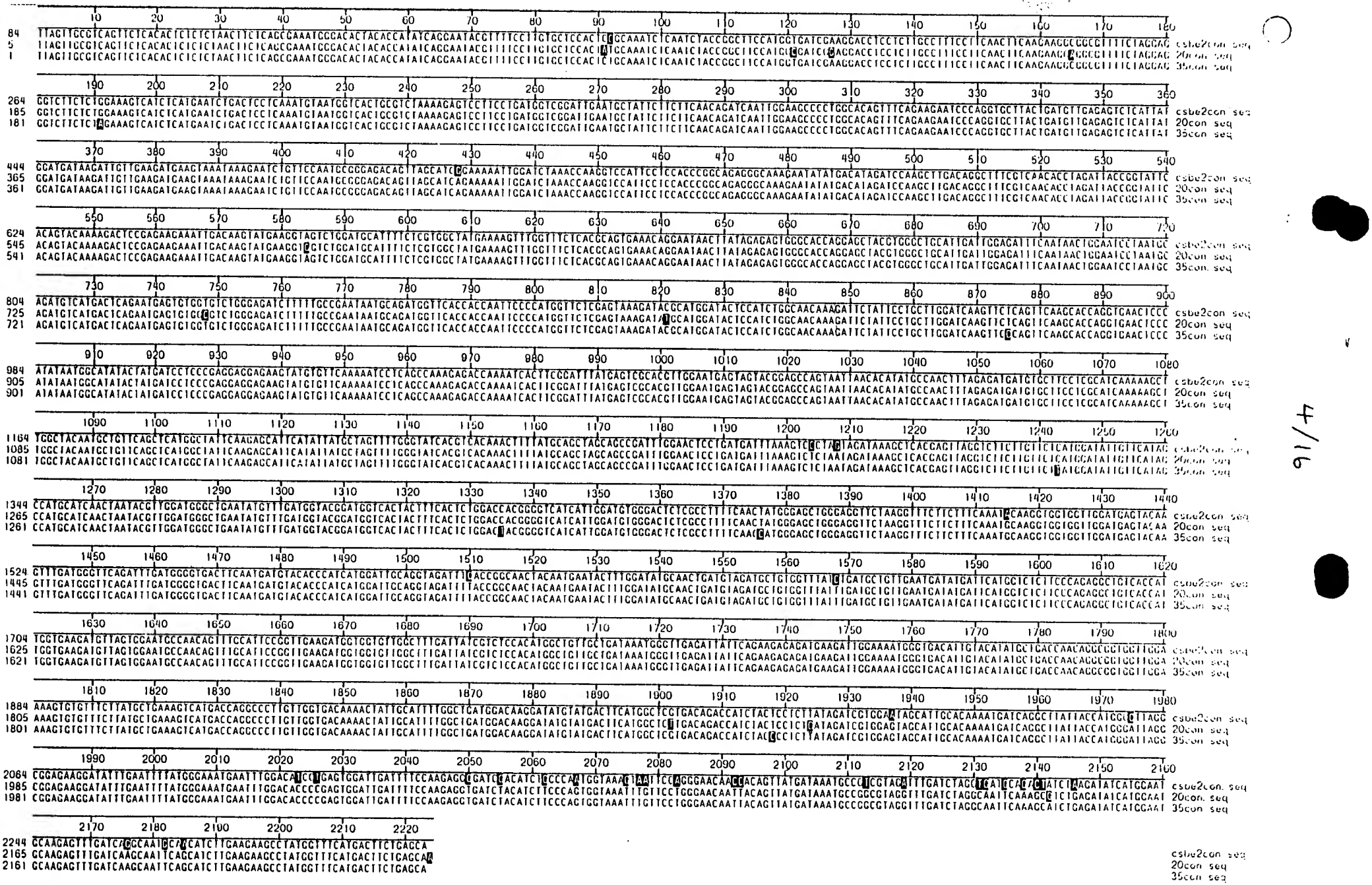


Fig. 4

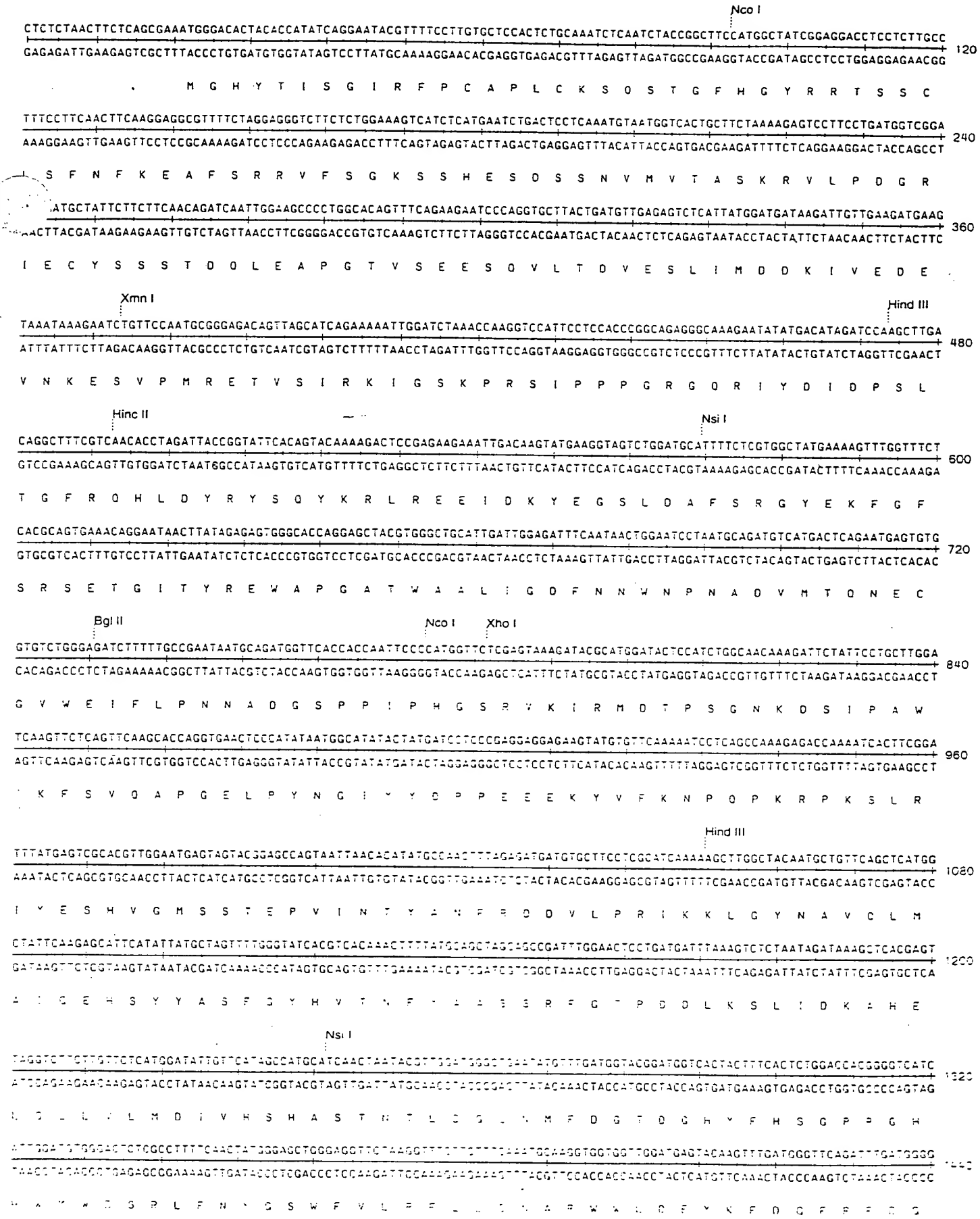

Nco I
120
M G H Y T I S G I R F P C A P L C K S Q S T G F H G Y R R T S S C
240
S F N F K E A F S R R V F S G K S S H E S Q S S N V M V T A S K R V L P D G R
Xmn I
Hind III
360
I E C Y S S S T D Q L E A P G T V S E E S Q V L T D V E S L I M D D K I V E D E
480
V N K E S V P M R E T V S I R K I G S K P R S I P P P G R G Q R I Y D I D P S L
Hinc II
Nsi I
600
T G F R Q H L D Y R Y S Q Y K R L R E E I D K Y E G S L Q A F S R G Y E K F G F
720
S R S E T G I T Y R E W A P G A T W A A L I G Q F N N W N P N A D V M T Q N E C
Bgl II
Nco I
Xho I
840
G V W E I F L P N N A D G S P P I P H G S R Y K I R M D T P S G N K D S I P A W
960
K F S V Q A P G E L P Y N G I Y Y Q P P E E E K Y V F K N P Q P K R P K S L R
Hind III
1080
I Y E S H V G M S S T E P V I N T Y A N F R D D V L P R I K K L G Y N A V C L M
1200
A I C E H S Y Y A S F G Y H V T N F T A A S G R F G T P D D L K S L I D K A H E
Nsi I
1320
N Q L L Y L M D I V H S H A S T N T L G G L N M F D G T D G H Y F H S G P R G H
1440

Fig. 4 (con.)

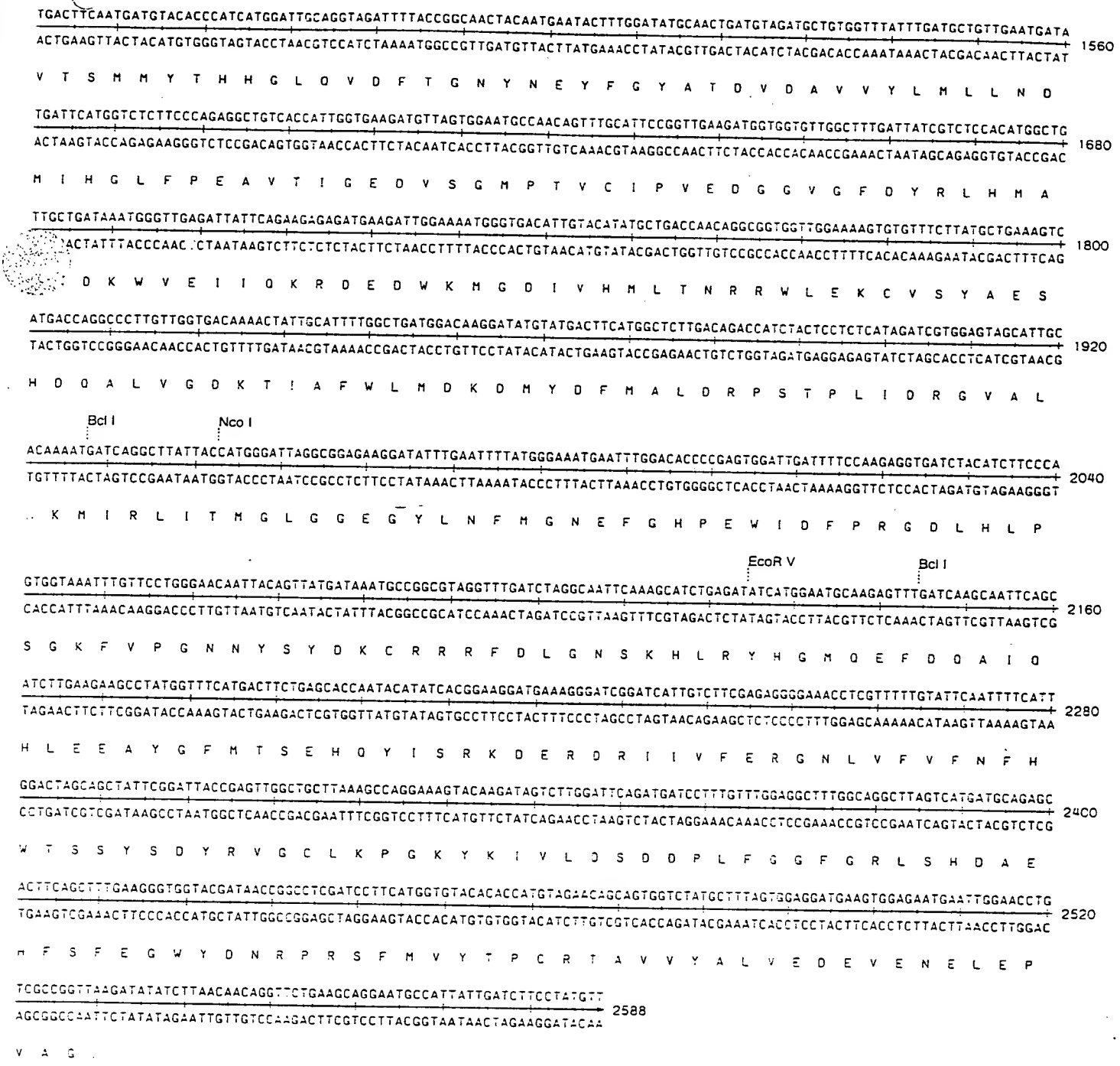

TGACTTCAATGATGTACACCCATCATGGATTGCAGGTAGATTTTACCGGCAACTACAATGAATACTTTGGATATGCAACTGATGTAGATGCTGTGGTTTATTTGATGCTGTTGAATGATA
ACTGAAGTTACTACATGTGGGTAGTACCTAACGTCCATCTAAAATGGCCGTTGATGTTACTTATGAAACCTATACGTTGACTACATCTACGACACCAAATAAACTACGACAACTTACTAT
1560
V T S M M Y T H H G L Q V D F T G N Y N E Y F G Y A T D V D A V V Y L M L L N D
TGATTCATGGTCTCTTCCCAGAGGCTGTCACCATTGGTGAAGATGTTAGTGGAATGCCAACAGTTTGCATTCCGGTTGAAGATGGTGGTGTTGGCTTTGATTATCGTCTCCACATGGCTG
ACTAAGTACCAGAGAAGGGTCTCCGACAGTGGTAACCACTTCTACAATCACCTTACGGTTGTCAAACGTAAGGCCAACTTCTACCACCACAACCGAAACTAATAGCAGAGGTGTACCGAC
1680
M I H G L F P E A V T I G E D V S G M P T V C I P V E D G G V G F D Y R L H M A
TTGCTGATAAATGGGTTGAGATTATTCAGAAGAGAGATGAAGATTGGAAAATGGGTGACATTGTACATATGCTGACCAACAGGCGGTGGTTGGAAAAGTGTGTTTCTTATGCTGAAAGTC
ACTATTTACCCAAC CTAATAAGTCTTCTCTCTACTTCTAACCTTTTACCCACTGTAACATGTATACGACTGGTTGTCCGCCACCAACCTTTTCACACAAAGAATACGACTTTCAG
1800
D K W V E I I Q K R D E D W K M G D I V H M L T N R R W L E K C V S Y A E S
ATGACCAGGCCCTTGTTGGTGACAAAACTATTGCATTTTGGCTGATGGACAAGGATATGTATGACTTCATGGCTCTTGACAGACCATCTACTCCTCTCATAGATCGTGGAGTAGCATTGC
TACTGGTCCGGGAACAACCACTGTTTTGATAACGTAAAACCGACTACCTGTTCCTATACATACTGAAGTACCGAGAACTGTCTGGTAGATGAGGAGAGTATCTAGCACCTCATCGTAACG
1920
H D Q A L V G D K T I A F W L M D K D M Y D F M A L D R P S T P L I D R G V A L
Bcl I
Nco I
ACAAAATGATCAGGCTTATTACCATGGGATTAGGCGGAGAAGGATATTTGAATTTTATGGGAAATGAATTTGGACACCCCGAGTGGATTGATTTTCCAAGAGGTGATCTACATCTTCCCA
TGTTTTACTAGTCCGAATAATGGTACCCTAATCCGCCTCTTCCTATAAACTTAAAATACCCTTTACTTAAACCTGTGGGGCTCACCTAACTAAAAGGTTCTCCACTAGATGTAGAAGGGT
2040
K M I R L I T M G L G G E G Y L N F M G N E F G H P E W I D F P R G D L H L P
EcoR V
Bcl I
GTGGTAAATTTGTTCCTGGGAACAATTACAGTTATGATAAATGCCGGCGTAGGTTTGATCTAGGCAATTCAAAGCATCTGAGATATCATGGAATGCAAGAGTTTGATCAAGCAATTCAGC
CACCATTTAAACAAGGACCCTTGTTAATGTCAATACTATTTACGGCCGCATCCAAACTAGATCCGTTAAGTTTCGTAGACTCTATAGTACCTTACGTTCTCAAACTAGTTCGTTAAGTCG
2160
S G K F V P G N N Y S Y D K C R R R F D L G N S K H L R Y H G M Q E F D Q A I Q
ATCTTGAAGAAGCCTATGGTTTCATGACTTCTGAGCACCAATACATATCACGGAAGGATGAAAGGGATCGGATCATTGTCTTCGAGAGGGGAAACCTCGTTTTTGTATTCAATTTTCATT
TAGAACTTCTTCGGATACCAAAGTACTGAAGACTCGTGGTTATGTATAGTGCCTTCCTACTTTCCCTAGCCTAGTAACAGAAGCTCTCCCCTTTGGAGCAAAAACATAAGTTAAAAGTAA
2280
H L E E A Y G F M T S E H Q Y I S R K D E R D R I I V F E R G N L V F V F N F H
GGACTAGCAGCTATTCGGATTACCGAGTTGGCTGCTTAAAGCCAGGAAAGTACAAGATAGTCTTGGATTCAGATGATCCTTTGTTTGGAGGCTTTGGCAGGCTTAGTCATGATGCAGAGC
CCTGATCGTCGATAAGCCTAATGGCTCAACCGACGAATTTCGGTCCTTTCATGTTCTATCAGAACCTAAGTCTACTAGGAAACAAACCTCCGAAACCGTCCGAATCAGTACTACGTCTCG
2400
W T S S Y S D Y R V G C L K P G K Y K I V L D S D D P L F G G F G R L S H D A E
ACTTCAGCTTTGAAGGGTGGTACGATAACCGGCCTCGATCCTTCATGGTGTACACACCATGTAGAACAGCAGTGGTCTATGCTTTAGTGGAGGATGAAGTGGAGAATGAATTGGAACCTG
TGAAGTCGAAACTTCCCACCATGCTATTGGCCGGAGCTAGGAAGTACCACATGTGTGGTACATCTTGTCGTCACCAGATACGAAATCACCTCCTACTTCACCTCTTACTTAACCTTGGAC
2520
F S F E G W Y D N R P R S F M V Y T P C R T A V V Y A L V E D E V E N E L E P
TCGCCGGTTAAGATATATCTTAACAACAGGTTCTGAAGCAGGAATGCCATTATTGATCTTCCTATGTT
AGCGGCCAATTCTATATAGAATTGTTGTCCAAGACTTCGTCCTTACGGTAATAACTAGAAGGATACAA
2588
V A G

Fig. 5

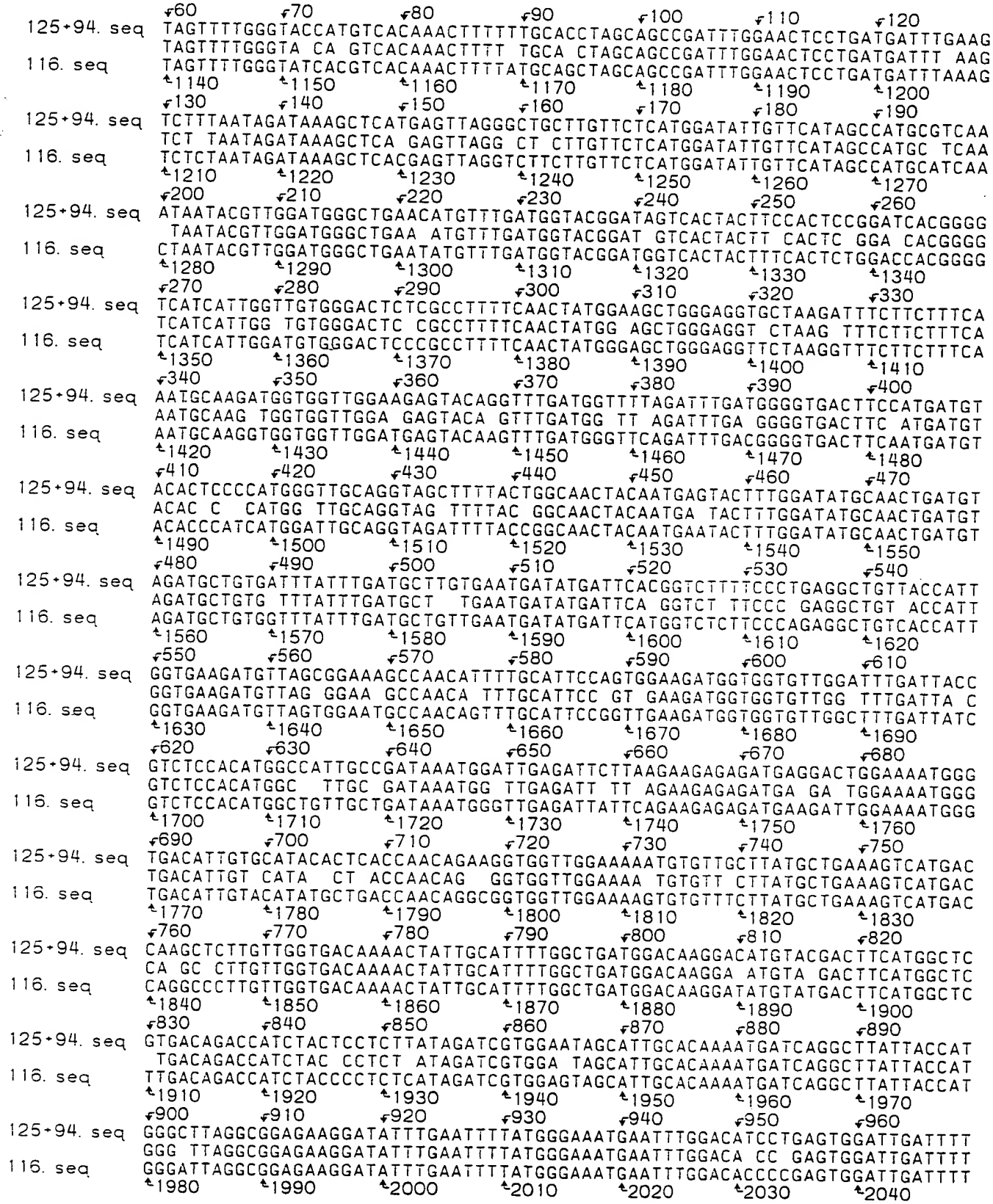

60 70 80 90 100 110 120
125+94. seq TAGTTTTGGGTACCATGTCACAAACTTTTTTGCACCTAGCAGCCGATTTGGAACTCCTGATGATTTGAAG
TAGTTTTGGGTA CA GTCACAAACTTTT TGCA CTAGCAGCCGATTTGGAACTCCTGATGATTT AAG
116. seq TAGTTTTGGGTATCACGTCACAAACTTTTATGCAGCTAGCAGCCGATTTGGAACTCCTGATGATTTAAAG
1140 1150 1160 1170 1180 1190 1200
130 140 150 160 170 180 190
125+94. seq TCTTTAATAGATAAAGCTCATGAGTTAGGGCTGCTTGTTCTCATGGATATTGTTCATAGCCATGCGTCAA
TCT TAATAGATAAAGCTCA GAGTTAGG CT CTTGTTCTCATGGATATTGTTCATAGCCATGC TCAA
116. seq TCTCTAATAGATAAAGCTCACGAGTTAGGTCTTCTTGTTCTCATGGATATTGTTCATAGCCATGCATCAA
1210 1220 1230 1240 1250 1260 1270
200 210 220 230 240 250 260
125+94. seq ATAATACGTTGGATGGGCTGAACATGTTTGATGGTACGGATAGTCACTACTTCCACTCCGGATCACGGGG
TAATACGTTGGATGGGCTGAA ATGTTTGATGGTACGGAT GTCACTACTT CACTC GGA CACGGGG
116. seq CTAATACGTTGGATGGGCTGAATATGTTTGATGGTACGGATGGTCACTACTTTCACTCTGGACCACGGGG
1280 1290 1300 1310 1320 1330 1340
270 280 290 300 310 320 330
125+94. seq TCATCATTGGTTGTGGGACTCTCGCCTTTTCAACTATGGAAGCTGGGAGGTGCTAAGATTTCTTCTTTCA
TCATCATTGG TGTGGGACTC CGCCTTTTCAACTATGG AGCTGGGAGGT CTAAG TTTCTTCTTTCA
116. seq TCATCATTGGATGTGGGACTCCCGCCTTTTCAACTATGGGAGCTGGGAGGTTCTAAGGTTTCTTCTTTCA
1350 1360 1370 1380 1390 1400 1410
340 350 360 370 380 390 400
125+94. seq AATGCAAGATGGTGGTTGGAAGAGTACAGGTTTGATGGTTTTAGATTTGATGGGGTGACTTCCATGATGT
AATGCAAG TGGTGGTTGGA GAGTACA GTTTGATGG TT AGATTTGA GGGGTGACTTC ATGATGT
116. seq AATGCAAGGTGGTGGTTGGATGAGTACAAGTTTGATGGGTTCAGATTTGACGGGGTGACTTCAATGATGT
1420 1430 1440 1450 1460 1470 1480
410 420 430 440 450 460 470
125+94. seq ACACTCCCCATGGGTTGCAGGTAGCTTTTACTGGCAACTACAATGAGTACTTTGGATATGCAACTGATGT
ACAC C CATGG TTGCAGGTAG TTTTAC GGCAACTACAATGA TACTTTGGATATGCAACTGATGT
116. seq ACACCCATCATGGATTGCAGGTAGATTTTACCGGCAACTACAATGAATACTTTGGATATGCAACTGATGT
1490 1500 1510 1520 1530 1540 1550
480 490 500 510 520 530 540
125+94. seq AGATGCTGTGATTTATTTGATGCTTGTGAATGATATGATTCACGGTCTTTTCCCTGAGGCTGTTACCATT
AGATGCTGTG TTTATTTGATGCT TGAATGATATGATTCA GGTCT TTCCC GAGGCTGT ACCATT
116. seq AGATGCTGTGGTTTTATTTGATGCTGTTGAATGATATGATTCATGGTCTCTTCCCAGAGGCTGTCACCATT
1560 1570 1580 1590 1600 1610 1620
550 560 570 580 590 600 610
125+94. seq GGTGAAGATGTTAGCGGAAAGCCAACATTTTGCATTCCAGTGGAAGATGGTGGTGTTGGATTTGATTACC
GGTGAAGATGTTAG GGAA GCCAACA TTTGCATTCC GT GAAGATGGTGGTGTTGG TTTGATTA C
116. seq GGTGAAGATGTTAGTGGAATGCCAACAGTTTGCATTCCGGTTGAAGATGGTGGTGTTGGCTTTGATTATC
1630 1640 1650 1660 1670 1680 1690
620 630 640 650 660 670 680
125+94. seq GTCTCCACATGGCCATTGCCGATAAATGGATTGAGATTCTTAAGAAGAGAGATGAGGACTGGAAAATGGG
GTCTCCACATGGC TTGC GATAAATGG TTGAGATT TT AGAAGAGAGATGA GA TGGAAAATGGG
116. seq GTCTCCACATGGCTGTTGCTGATAAATGGGTTGAGATTATTCAGAAGAGAGATGAAGATTGGAAAATGGG
1700 1710 1720 1730 1740 1750 1760
690 700 710 720 730 740 750
125+94. seq TGACATTGTGCATACACTCACCAACAGAAGGTGGTTGGAAAAATGTGTTGCTTATGCTGAAAGTCATGAC
TGACATTGT CATA CT ACCAACAG GGTGGTTGGAAAA TGTGTT CTTATGCTGAAAGTCATGAC
116. seq TGACATTGTACATATGCTGACCAACAGGCGGTGGTTGGAAAAGTGTGTTTCTTATGCTGAAAGTCATGAC
1770 1780 1790 1800 1810 1820 1830
760 770 780 790 800 810 820
125+94. seq CAAGCTCTTGTTGGTGACAAAACTATTGCATTTTGGCTGATGGACAAGGACATGTACGACTTCATGGCTC
CA GC CTTGTTGGTGACAAAACTATTGCATTTTGGCTGATGGACAAGGA ATGTA GACTTCATGGCTC
116. seq CAGGCCCTTGTTGGTGACAAAACTATTGCATTTTGGCTGATGGACAAGGATATGTATGACTTCATGGCTC
1840 1850 1860 1870 1880 1890 1900
830 840 850 860 870 880 890
125+94. seq GTGACAGACCATCTACTCCTCTTATAGATCGTGGAATAGCATTGCACAAAATGATCAGGCTTATTACCAT
TGACAGACCATCTAC CCTCT ATAGATCGTGGA TAGCATTGCACAAAATGATCAGGCTTATTACCAT
116. seq TTGACAGACCATCTACCCCTCTCATAGATCGTGGAGTAGCATTGCACAAAATGATCAGGCTTATTACCAT
1910 1920 1930 1940 1950 1960 1970
900 910 920 930 940 950 960
125+94. seq GGGCTTAGGCGGAGAAGGATATTTGAATTTTATGGGAAATGAATTTGGACATCCTGAGTGGATTGATTTT
GGG TTAGGCGGAGAAGGATATTTGAATTTTATGGGAAATGAATTTGGACA CC GAGTGGATTGATTTT
116. seq GGGATTAGGCGGAGAAGGATATTTGAATTTTATGGGAAATGAATTTGGACACCCCGAGTGGATTGATTTT
1980 1990 2000 2010 2020 2030 2040

Fig. 5 (con)


```
                 ←970       ←980       ←990       ←1000      ←1010      ←1020      ←1030
125+94. seq      CCAAGAGGGGATCGACATCTGCCCAATGGTAAAGTAATTCCAGGGAACAACCACAGTTATGATAAATGCC
                 CCAAGAGG GATC ACATCT CCCA TGGTAAA T  TTCC GGGAACAA  ACAGTTATGATAAATGCC
116. seq         CCAAGAGGTGATCTACATCTTCCCAGTGGTAAATTTGTTCCTGGGAACAATTACAGTTATGATAAATGCC
                 →2050      →2060      →2070      →2080      →2090      →2100      →2110
                 ←1040      ←1050      ←1060      ←1070      ←1080      ←1090      ←1100
125+94. seq      GTCGTAGATTTGATCTAGGTGATGCAGACTATCTAAGATATCATGGAATGCAAGAGTTTGATCAGGCAAT
                 G CGTAG TTTGATCTAGG  AT CA A  ATCT AGATATCATGGAATGCAAGAGTTTGATCA GCAAT
116. seq         GGCGTAGGTTTGATCTAGGCAATTCAAAGCATCTGAGATATCATGGAATGCAAGAGTTTGATCAAGCAAT
                 →2120      →2130      →2140      →2150      →2160      →2170      →2180
                 ←1110      ←1120      ←1130      ←1140      ←1150      ←1160      ←1170
125+94. seq      GCAACATCTTGAAGAAGCCTATGGTTTCATGACTTCTGAGCACCAGTATATATCACGGAAGGATGAAGGA
                  CA CATCTTGAAGAAGCCTATGGTTTCATGACTTCTGAGCACCA TA ATATCACGGAAGGATGAA G
116. seq         TCAGCATCTTGAAGAAGCCTATGGTTTCATGACTTCTGAGCACCAATACATATCACGGAAGGATGAAAGG
                 →2190      →2200      →2210      →2220      →2230      →2240      →2250
                 ←1180      ←1190      ←1200      ←1210      ←1220      ←1230      ←1240
125+94. seq      GATCGGATCATTGTCTTTGAGAGGGGAAACCTTGTTTTTGTATTCAACTTTCATTGGACTAACAGCTATT
                 GATCGGATCATTGTCTT GAGAGGGGAAACCT GTTTTTGTATTCAA TTTCATTGGACTA CAGCTATT
116. seq         GATCGGATCATTGTCTTCGAGAGGGGAAACCTCGTTTTTGTATTCAATTTTCATTGGACTAGCAGCTATT
                 →2260      →2270      →2280      →2290      →2300      →2310      →2320
                 ←1250      ←1260      ←1270      ←1280      ←1290      ←1300      ←1310
125+94. seq      CAGATTACCGAGTTGGCTGCTTCAAGTCAGGAAAGTACAAGATTGTTTTGGACTCGGATGATGGCTTGTT
                 C GATTACCGAGTTGGCTGCTT AAG CAGGAAAGTACAAGAT GT TTGGA TC GATGAT  TTGTT
116. seq         CGGATTACCGAGTTGGCTGCTTAAAGCCAGGAAAGTACAAGATAGTCTTGGATTCAGATGATCCTTTGTT
                 →2330      →2340      →2350      →2360      →2370      →2380      →2390
                 ←1320      ←1330      ←1340      ←1350      ←1360      ←1370      ←1380
125+94. seq      TGGAGGCTTCAACAGGCTTAGTCATGATGCCGAGCACTTCACCTTTGACGGGTGGTATGATAACCGGCCT
                 TGGAGGCTT   CAGGCTTAGTCATGATGC GAGCACTTCA CTTTGA GGGTGGTA GATAACCGGCCT
116. seq         TGGAGGCTTTGGCAGGCTTAGTCATGATGCAGAGCACTTCAGCTTTGAAGGGTGGTACGATAACCGGCCT
                 →2400      →2410      →2420      →2430      →2440      →2450      →2460
                 ←1390      ←1400      ←1410      ←1420      ←1430      ←1440      ←1450
125+94. seq      CGGTCCTTCATGGTATATGCACCATCTAGGACAGCAGTGGTCCATGCTTTAGTAGAAGATGAAG
                 CG TCCTTCATGGT TA  CACCAT TAG ACAGCAGTGGTC ATGCTTTAGT GA GATGAAG
116. seq         CGATCCTTCATGGTGTACACACCATGTAGAACAGCAGTGGTCTATGCTTTAGTGGAGGATGAAG
                 →2470      →2480      →2490      →2500      →2510      →2520      →2530
```

Fig. 6


```
                  ↓10        ↓20        ↓30        ↓40        ↓50        ↓60        ↓70
125-94. pro SFGYHVTNFFAPSSRFGTPDDLKSLIDKAHELGLLVLMDIVHSHASNNTLDGLNMFDGTDSHYFHSGSRG
            SFGYHVTNF: A: SSRFGTPDDLKSLIDKAHELGLLVLMDIVHSHAS. NTLDGLNMFDGTD: HYFHSG: RG
116. pro    SFGYHVTNFYAASSRFGTPDDLKSLIDKAHELGLLVLMDIVHSHASTNTLDGLNMFDGTDGHYFHSGPRG
                  ↑370       ↑380       ↑390       ↑400       ↑410       ↑420       ↑430
                  ↓80        ↓90        ↓100       ↓110       ↓120       ↓130       ↓140
125-94. pro HHWLWDSRLFNYGSWEVLRFLLSNARWWLEEYRFDGFRFDGVTSMMYTPHGLQVAFTGNYNEYFGYATDV
            HHW: WDSRLFNYGSWEVLRFLLSNARWWL: EY: FDGFRFDGVTSMMYT. HGLQV. FTGNYNEYFGYATDV
116. pro    HHWMWDSRLFNYGSWEVLRFLLSNARWWLDEYKFDGFRFDGVTSMMYTHHGLQVDFTGNYNEYFGYATDV
                  ↑440       ↑450       ↑460       ↑470       ↑480       ↑490       ↑500
                  ↓150       ↓160       ↓170       ↓180       ↓190       ↓200       ↓210
125-94. pro DAVIYLMLVNDMIHGLFPEAVTIGEDVSGKPTFCIPVEDGGVGFDYRLHMAIADKWIEILKKRDEDWKMG
            DAV: YLML: NDMIHGLFPEAVTIGEDVSG. PT CIPVEDGGVGFDYRLHMA: ADKW: EI: : KRDEDWKMG
116. pro    DAVVYLMLLNDMIHGLFPEAVTIGEDVSGMPTVCIPVEDGGVGFDYRLHMAVADKWVEIIQKRDEDWKMG
                  ↑510       ↑520       ↑530       ↑540       ↑550       ↑560       ↑570
                  ↓220       ↓230       ↓240       ↓250       ↓260       ↓270       ↓280
125-94. pro DIVHTLTNRRWLEKCVAYAESHDQALVGDKTIAFWLMDKDMYDFMARDRPSTPLIDRGIALHKMIRLITM
            DIVH LTNRRWLEKCV: YAESHDQALVGDKTIAFWLMDKDMYDFMA DRPSTPLIDRG: ALHKMIRLITM
116. pro    DIVHMLTNRRWLEKCVSYAESHDQALVGDKTIAFWLMDKDMYDFMALDRPSTPLIDRGVALHKMIRLITM
                  ↑580       ↑590       ↑600       ↑610       ↑620       ↑630       ↑640
                  ↓290       ↓300       ↓310       ↓320       ↓330       ↓340       ↓350
125-94. pro GLGGEGYLNFMGNEFGHPEWIDFPRGDRHLPNGKVIPGNNHSYDKCRRRFDLGDADYLRYHGMQEFDQAM
            GLGGEGYLNFMGNEFGHPEWIDFPRGD HLP: GK : PGNN. SYDKCRRRFDLG: : . . LRYHGMQEFDQA:
116. pro    GLGGEGYLNFMGNEFGHPEWIDFPRGDLHLPSGKFVPGNNYSYDKCRRRFDLGNSKHLRYHGMQEFDQAI
                  ↑650       ↑660       ↑670       ↑680       ↑690       ↑700       ↑710
                  ↓360       ↓370       ↓380       ↓390       ↓400       ↓410       ↓420
125-94. pro QHLEEAYGFMTSEHQYISRKDEGDRIIVFERGNLVFVFNFHWTNSYSDYRVGCFKSGKYKIVLDSDDGLF
            QHLEEAYGFMTSEHQYISRKDE DRIIVFERGNLVFVFNFHWT: SYSDYRVGC: K: GKYKIVLDSDD LF
116. pro    QHLEEAYGFMTSEHQYISRKDERDRIIVFERGNLVFVFNFHWTSSYSDYRVGCLKPGKYKIVLDSDDPLF
                  ↑720       ↑730       ↑740       ↑750       ↑760       ↑770       ↑780
                  ↓430       ↓440       ↓450       ↓460       ↓470
125-94. pro GGFNRLSHDAEHFTFDGWYDNRPRSFMVYAPSRTAVVHALVEDEENEAENEVES
            GGF. RLSHDAEHF: F: GWYDNRPRSFMVY: P. RTAVV. ALVEDE : : . : V. :
116. pro    GGFGRLSHDAEHFSFEGWYDNRPRSFMVYTPCRTAVVYALVEDEVENEVEPVAG
                  ↑790       ↑800       ↑810       ↑820       ↑830
```

Fig. 7

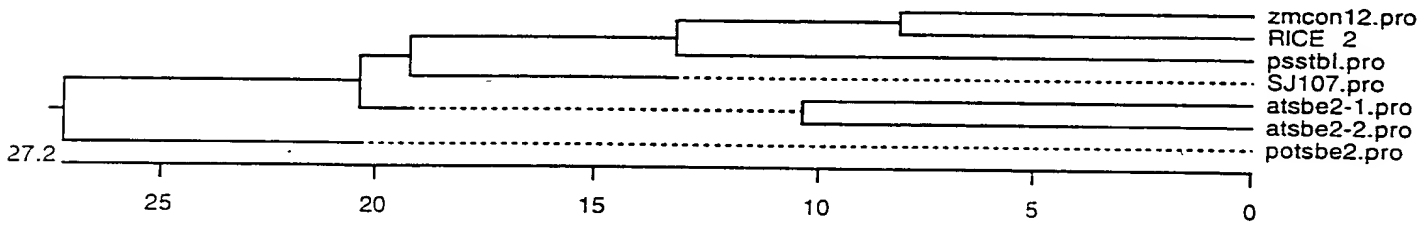
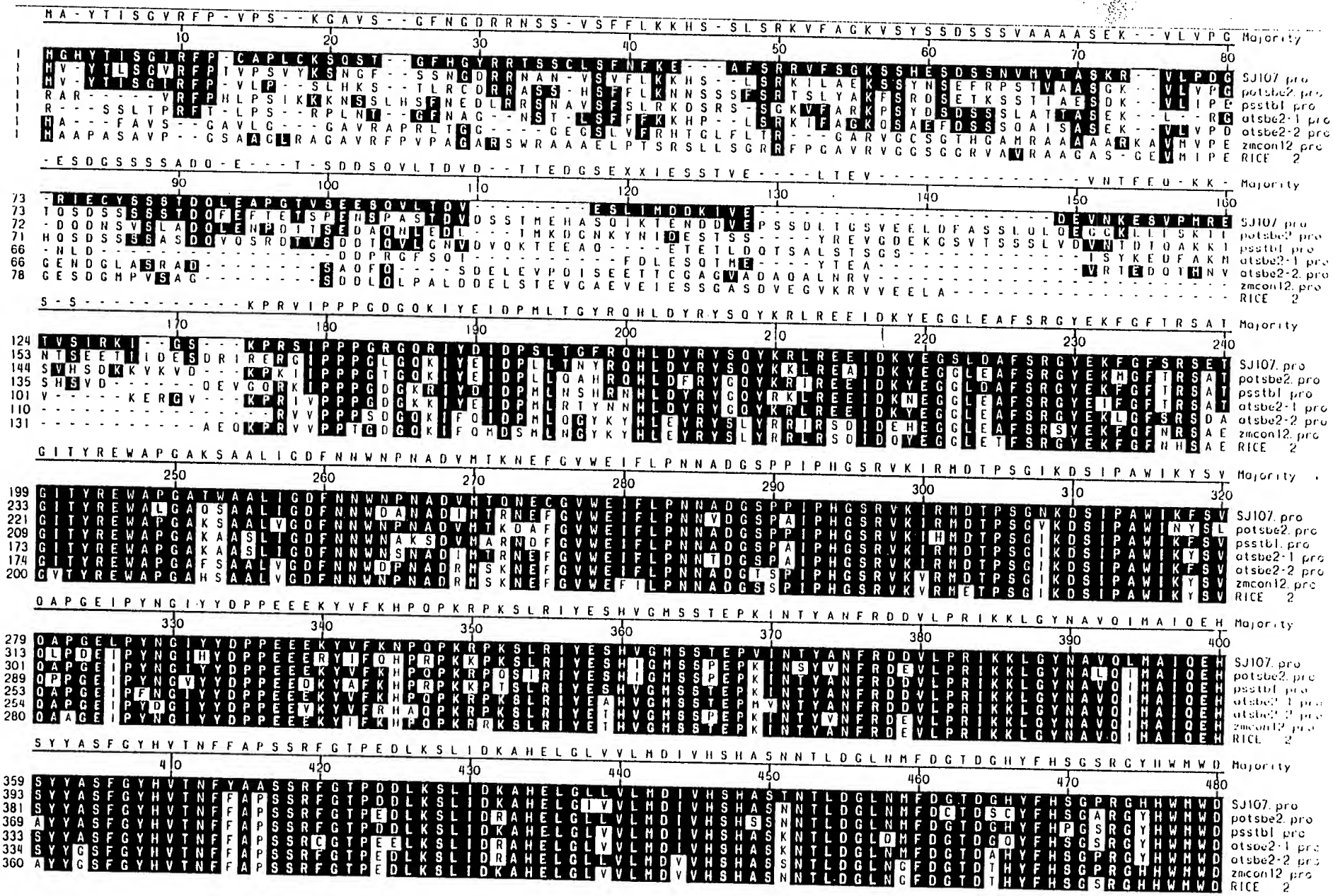
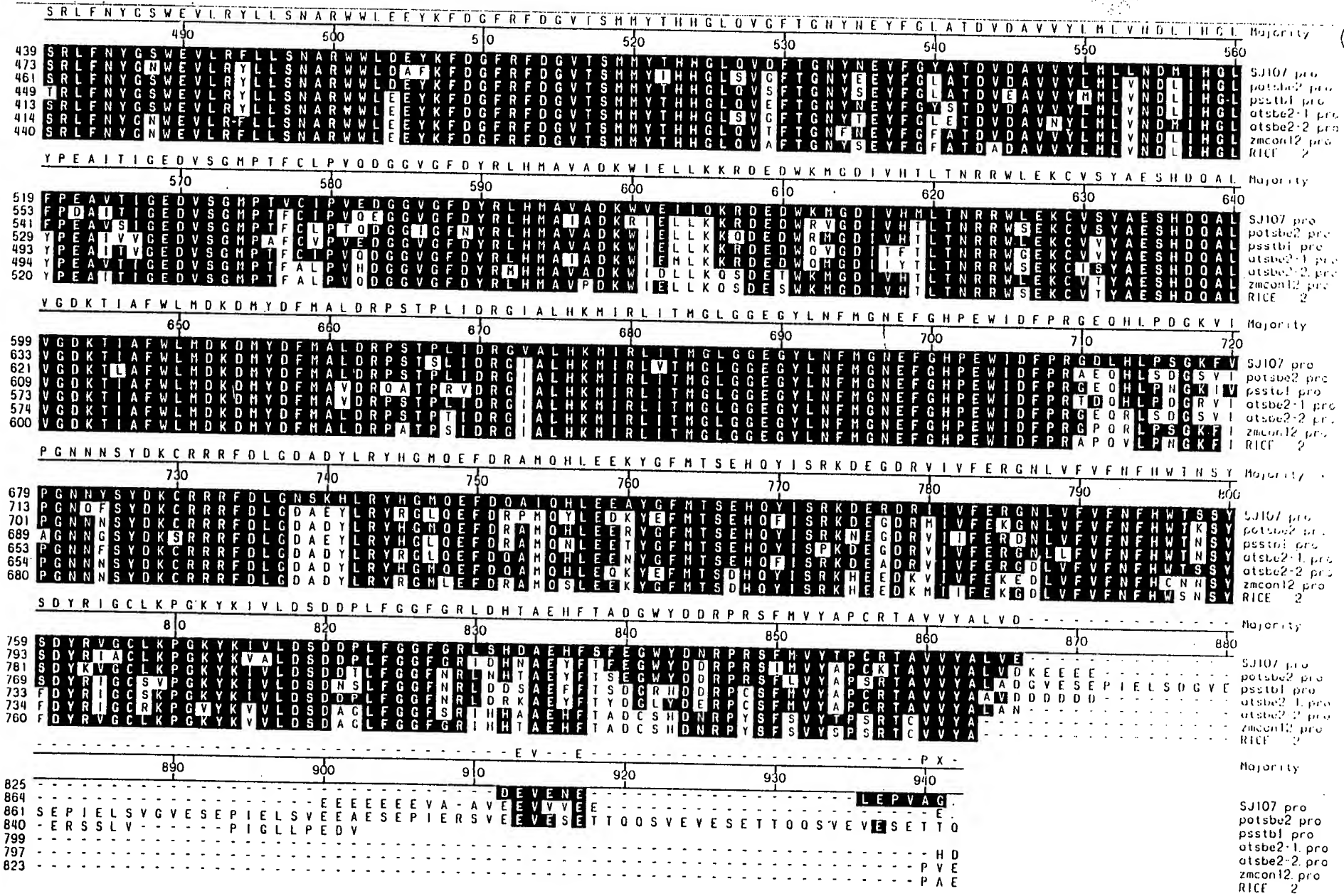

zmcon12.pro
RICE 2
psstbl.pro
SJ107.pro
atsbe2-1.pro
atsbe2-2.pro
potsbe2.pro
27.2
25
20
15
10
5
0

Fig. 8


Majority
SJ107.pro
potsbe2.pro
psstbl pro
atsbe2-1 pro
atsbe2-2 pro
zmcon12.pro
RICE 2

11/16

Fig 8 (con.)



12/16

Fig. 9

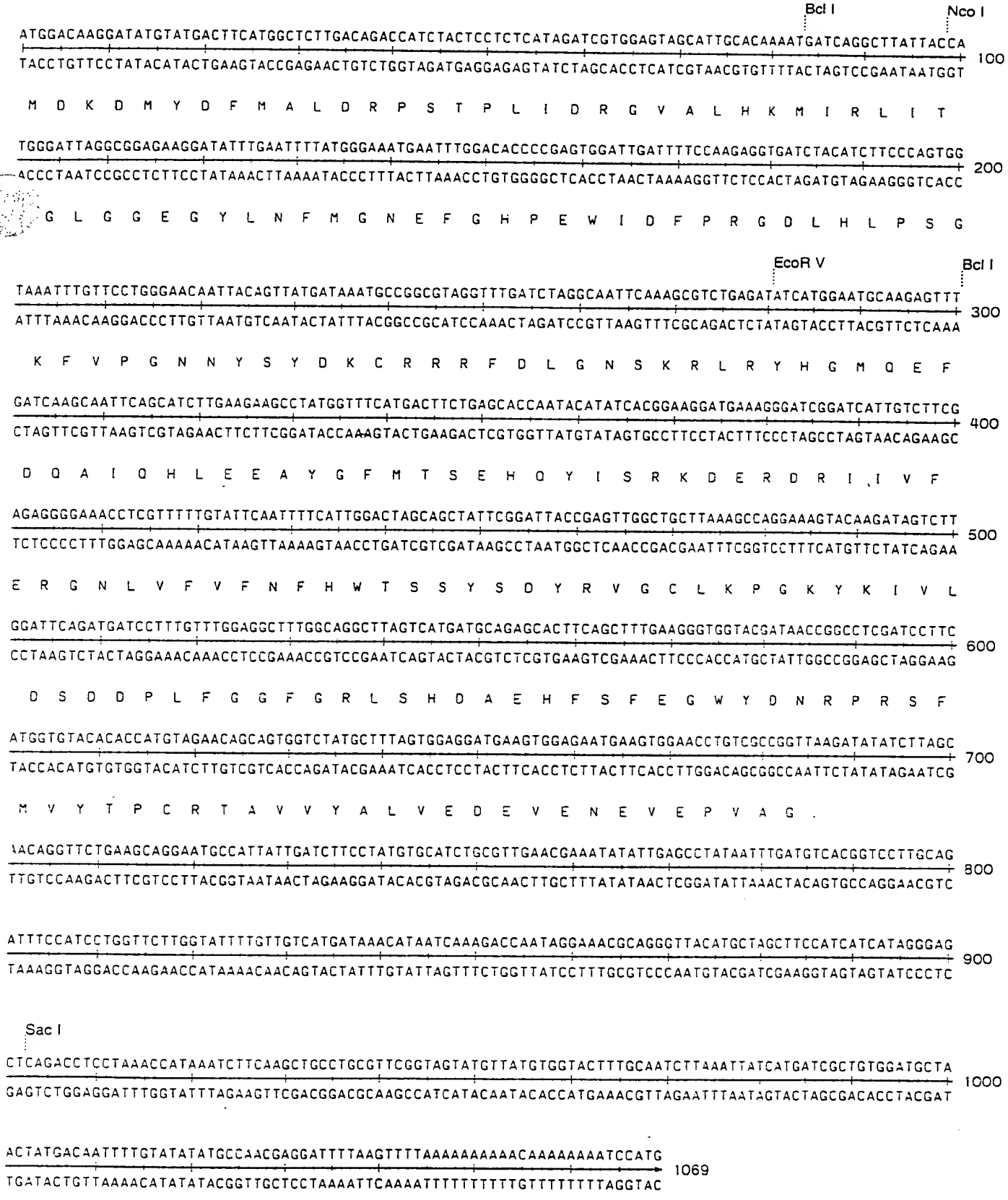

Bcl I
Nco I
ATGGACAAGGATATGTATGACTTCATGGCTCTTGACAGACCATCTACTCCTCTCATAGATCGTGGAGTAGCATTGCACAAAATGATCAGGCTTATTACCA
TACCTGTTCCTATACATACTGAAGTACCGAGAACTGTCTGGTAGATGAGGAGAGTATCTAGCACCTCATCGTAACGTGTTTTACTAGTCCGAATAATGGT
100
M D K D M Y D F M A L D R P S T P L I D R G V A L H K M I R L I T
TGGGATTAGGCGGAGAAGGATATTTGAATTTTATGGGAAATGAATTTGGACACCCCGAGTGGATTGATTTTCCAAGAGGTGATCTACATCTTCCCAGTGG
ACCCTAATCCGCCTCTTCCTATAAACTTAAAATACCCTTTACTTAAACCTGTGGGGCTCACCTAACTAAAAGGTTCTCCACTAGATGTAGAAGGGTCACC
200
G L G G E G Y L N F M G N E F G H P E W I D F P R G D L H L P S G
EcoR V
Bcl I
TAAATTTGTTCCTGGGAACAATTACAGTTATGATAAATGCCGGCGTAGGTTTGATCTAGGCAATTCAAAGCGTCTGAGATATCATGGAATGCAAGAGTTT
ATTTAAACAAGGACCCTTGTTAATGTCAATACTATTTACGGCCGCATCCAAACTAGATCCGTTAAGTTTCGCAGACTCTATAGTACCTTACGTTCTCAAA
300
K F V P G N N Y S Y D K C R R R F D L G N S K R L R Y H G M Q E F
GATCAAGCAATTCAGCATCTTGAAGAAGCCTATGGTTTCATGACTTCTGAGCACCAATACATATCACGGAAGGATGAAAGGGATCGGATCATTGTCTTCG
CTAGTTCGTTAAGTCGTAGAACTTCTTCGGATACCAAAGTACTGAAGACTCGTGGTTATGTATAGTGCCTTCCTACTTTCCCTAGCCTAGTAACAGAAGC
400
D Q A I Q H L E E A Y G F M T S E H Q Y I S R K D E R D R I I V F
AGAGGGGAAACCTCGTTTTTGTATTCAATTTTCATTGGACTAGCAGCTATTCGGATTACCGAGTTGGCTGCTTAAAGCCAGGAAAGTACAAGATAGTCTT
TCTCCCCTTTGGAGCAAAAACATAAGTTAAAAGTAACCTGATCGTCGATAAGCCTAATGGCTCAACCGACGAATTTCGGTCCTTTCATGTTCTATCAGAA
500
E R G N L V F V F N F H W T S S Y S D Y R V G C L K P G K Y K I V L
GGATTCAGATGATCCTTTGTTTGGAGGCTTTGGCAGGCTTAGTCATGATGCAGAGCACTTCAGCTTTGAAGGGTGGTACGATAACCGGCCTCGATCCTTC
CCTAAGTCTACTAGGAAACAAACCTCCGAAACCGTCCGAATCAGTACTACGTCTCGTGAAGTCGAAACTTCCCACCATGCTATTGGCCGGAGCTAGGAAG
600
D S D D P L F G G F G R L S H D A E H F S F E G W Y D N R P R S F
ATGGTGTACACACCATGTAGAACAGCAGTGGTCTATGCTTTAGTGGAGGATGAAGTGGAGAATGAAGTGGAACCTGTCGCCGGTTAAGATATATCTTAGC
TACCACATGTGTGGTACATCTTGTCGTCACCAGATACGAAATCACCTCCTACTTCACCTCTTACTTCACCTTGGACAGCGGCCAATTCTATATAGAATCG
700
M V Y T P C R T A V V Y A L V E D E V E N E V E P V A G .
TACAGGTTCTGAAGCAGGAATGCCATTATTGATCTTCCTATGTGCATCTGCGTTGAACGAAATATATTGAGCCTATAATTTGATGTCACGGTCCTTGCAG
ATGTCCAAGACTTCGTCCTTACGGTAATAACTAGAAGGATACACGTAGACGCAACTTGCTTTATATAACTCGGATATTAAACTACAGTGCCAGGAACGTC
800
ATTTCCATCCTGGTTCTTGGTATTTTGTTGTCATGATAAACATAATCAAAGACCAATAGGAAACGCAGGGTTACATGCTAGCTTCCATCATCATAGGGAG
TAAAGGTAGGACCAAGAACCATAAAACAACAGTACTATTTGTATTAGTTTCTGGTTATCCTTTGCGTCCCAATGTACGATCGAAGGTAGTAGTATCCCTC
900
Sac I
CTCAGACCTCCTAAACCATAAATCTTCAAGCTGCCTGCGTTCGGTAGTATGTTATGTGGTACTTTGCAATCTTAAATTATCATGATCGCTGTGGATGCTA
GAGTCTGGAGGATTTGGTATTTAGAAGTTCGACGGACGCAAGCCATCATACAATACACCATGAAACGTTAGAATTTAATAGTACTAGCGACACCTACGAT
1000
ACTATGACAATTTTGTATATATGCCAACGAGGATTTTAAGTTTTAAAAAAAAAACAAAAAAAATCCATG
TGATACTGTTAAAACATATATACGGTTGCTCCTAAAATTCAAAATTTTTTTTTTGTTTTTTTTAGGTAC
1069

Fig. 10

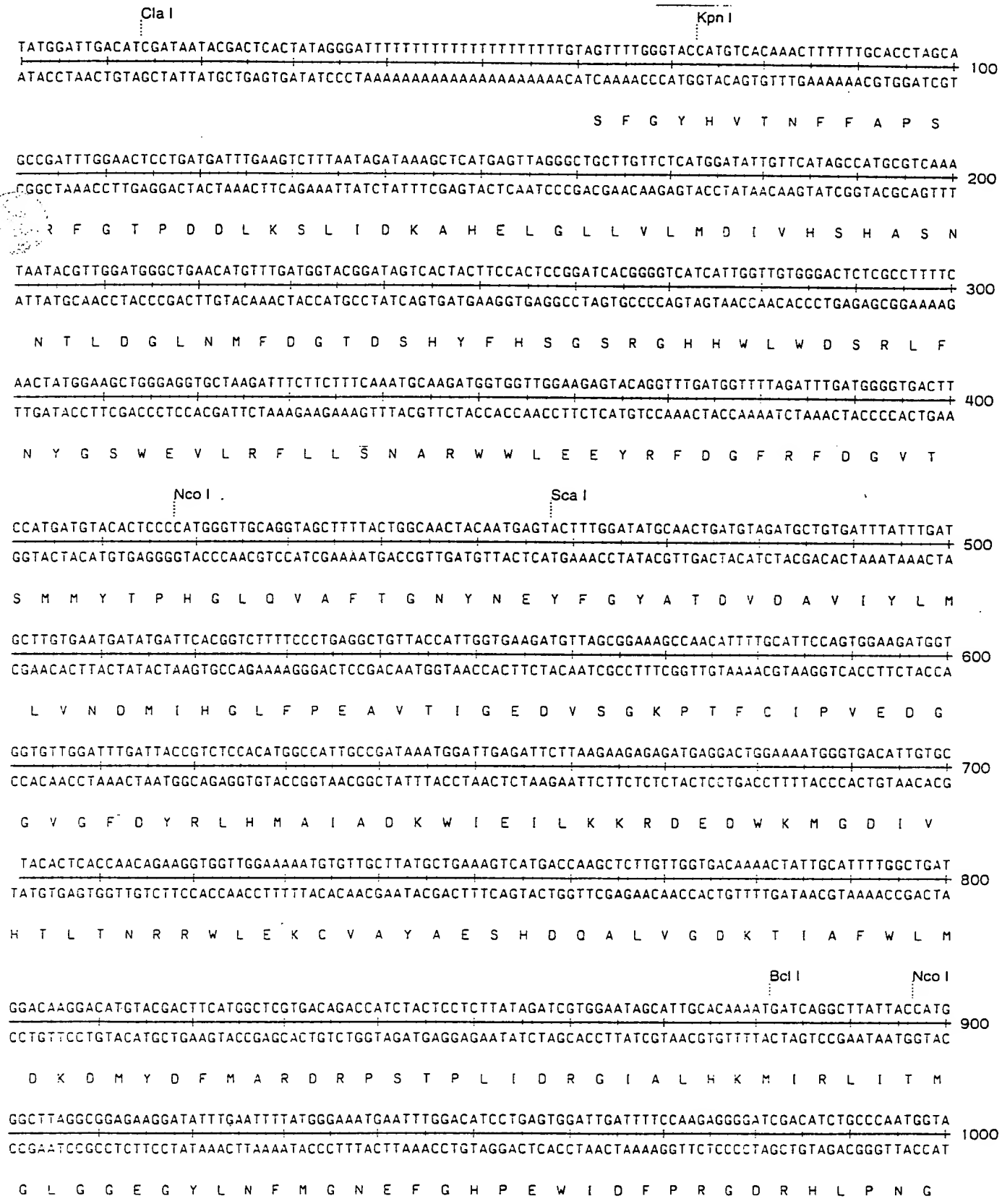

Cla I
Kpn I
TATGGATTGACATCGATAATACGACTCACTATAGGGATTTTTTTTTTTTTTTTTTTTTGTAGTTTTGGGTACCATGTCACAAACTTTTTTGCACCTAGCA
ATACCTAACTGTAGCTATTATGCTGAGTGATATCCCTAAAAAAAAAAAAAAAAAAAAAACATCAAAACCCATGGTACAGTGTTTGAAAAAACGTGGATCGT
100
S F G Y H V T N F F A P S
GCCGATTTGGAACTCCTGATGATTTGAAGTCTTTAATAGATAAAGCTCATGAGTTAGGGCTGCTTGTTCTCATGGATATTGTTCATAGCCATGCGTCAAA
CGGCTAAACCTTGAGGACTACTAAACTTCAGAAATTATCTATTTCGAGTACTCAATCCCGACGAACAAGAGTACCTATAACAAGTATCGGTACGCAGTTT
200
R F G T P D D L K S L I D K A H E L G L L V L M D I V H S H A S N
TAATACGTTGGATGGGCTGAACATGTTTGATGGTACGGATAGTCACTACTTCCACTCCGGATCACGGGGTCATCATTGGTTGTGGGACTCTCGCCTTTTC
ATTATGCAACCTACCCGACTTGTACAAACTACCATGCCTATCAGTGATGAAGGTGAGGCCTAGTGCCCCAGTAGTAACCAACACCCTGAGAGCGGAAAAG
300
N T L D G L N M F D G T D S H Y F H S G S R G H H W L W D S R L F
AACTATGGAAGCTGGGAGGTGCTAAGATTTCTTCTTTCAAATGCAAGATGGTGGTTGGAAGAGTACAGGTTTGATGGTTTTAGATTTGATGGGGTGACTT
TTGATACCTTCGACCCTCCACGATTCTAAAGAAGAAAGTTTACGTTCTACCACCAACCTTCTCATGTCCAAACTACCAAAATCTAAACTACCCCACTGAA
400
N Y G S W E V L R F L L S N A R W W L E E Y R F D G F R F D G V T
Nco I
Sca I
CCATGATGTACACTCCCCATGGGTTGCAGGTAGCTTTTACTGGCAACTACAATGAGTACTTTGGATATGCAACTGATGTAGATGCTGTGATTTATTTGAT
GGTACTACATGTGAGGGGTACCCAACGTCCATCGAAAATGACCGTTGATGTTACTCATGAAACCTATACGTTGACTACATCTACGACACTAAATAAACTA
500
S M M Y T P H G L Q V A F T G N Y N E Y F G Y A T D V D A V I Y L M
GCTTGTGAATGATATGATTCACGGTCTTTTCCCTGAGGCTGTTACCATTGGTGAAGATGTTAGCGGAAAGCCAACATTTTGCATTCCAGTGGAAGATGGT
CGAACACTTACTATACTAAGTGCCAGAAAAGGGACTCCGACAATGGTAACCACTTCTACAATCGCCTTTCGGTTGTAAAACGTAAGGTCACCTTCTACCA
600
L V N D M I H G L F P E A V T I G E D V S G K P T F C I P V E D G
GGTGTTGGATTTGATTACCGTCTCCACATGGCCATTGCCGATAAATGGATTGAGATTCTTAAGAAGAGAGATGAGGACTGGAAAATGGGTGACATTGTGC
CCACAACCTAAACTAATGGCAGAGGTGTACCGGTAACGGCTATTTACCTAACTCTAAGAATTCTTCTCTCTACTCCTGACCTTTTACCCACTGTAACACG
700
G V G F D Y R L H M A I A D K W I E I L K K R D E D W K M G D I V
TACACTCACCAACAGAAGGTGGTTGGAAAAATGTGTTGCTTATGCTGAAAGTCATGACCAAGCTCTTGTTGGTGACAAAACTATTGCATTTTGGCTGAT
TATGTGAGTGGTTGTCTTCCACCAACCTTTTTACACAACGAATACGACTTTCAGTACTGGTTCGAGAACAACCACTGTTTTGATAACGTAAAACCGACTA
800
H T L T N R R W L E K C V A Y A E S H D Q A L V G D K T I A F W L M
Bcl I
Nco I
GGACAAGGACATGTACGACTTCATGGCTCGTGACAGACCATCTACTCCTCTTATAGATCGTGGAATAGCATTGCACAAAATGATCAGGCTTATTACCATG
CCTGTTCCTGTACATGCTGAAGTACCGAGCACTGTCTGGTAGATGAGGAGAATATCTAGCACCTTATCGTAACGTGTTTTACTAGTCCGAATAATGGTAC
900
D K D M Y D F M A R D R P S T P L I D R G I A L H K M I R L I T M
GGCTTAGGCGGAGAAGGATATTTGAATTTTATGGGAAATGAATTTGGACATCCTGAGTGGATTGATTTTCCAAGAGGGGATCGACATCTGCCCAATGGTA
CCGAATCCGCCTCTTCCTATAAACTTAAAATACCCTTTACTTAAACCTGTAGGACTCACCTAACTAAAAGGTTCTCCCCTAGCTGTAGACGGGTTACCAT
1000
G L G G E G Y L N F M G N E F G H P E W I D F P R G D R H L P N G

Fig. 10 (con.)

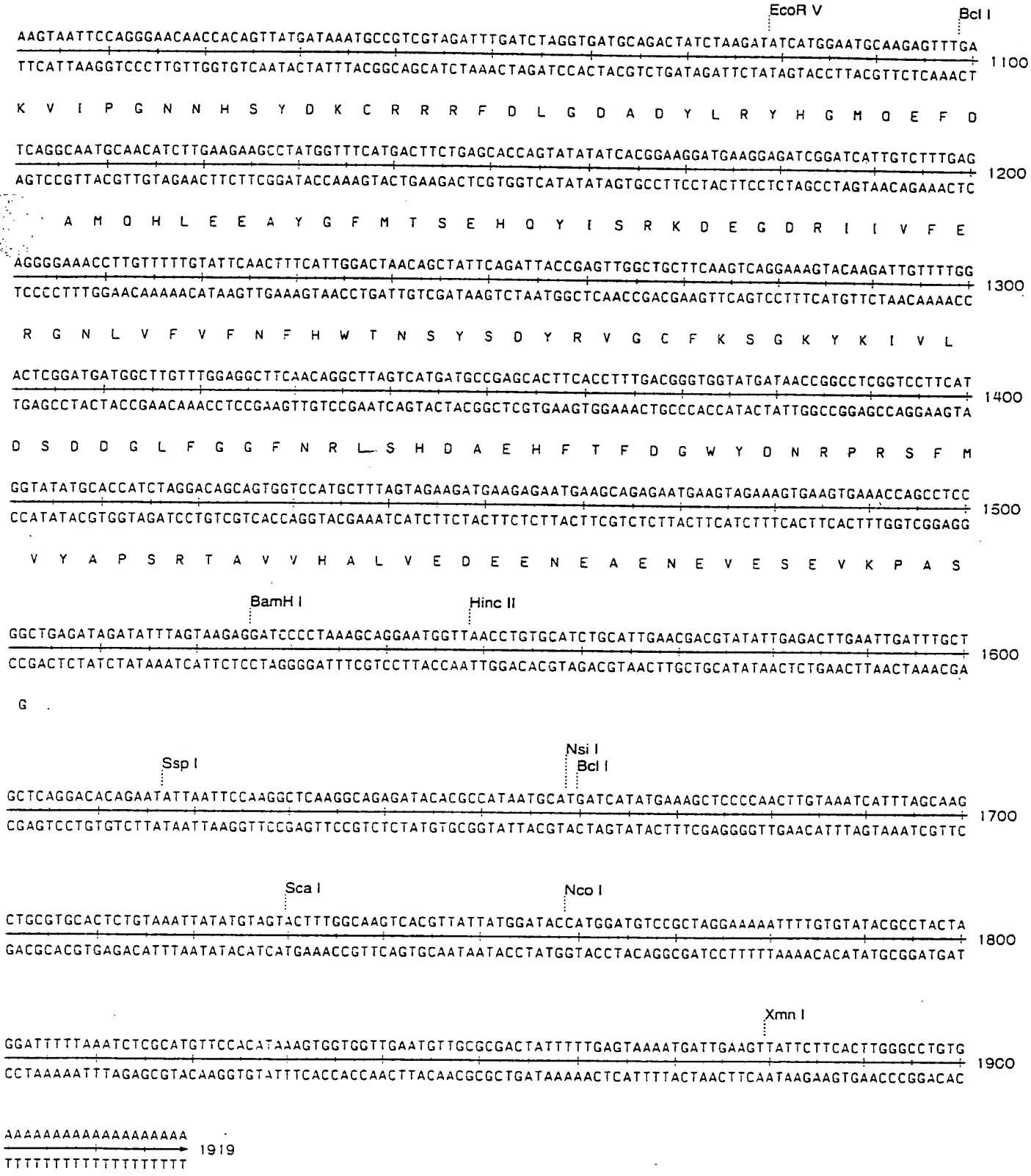
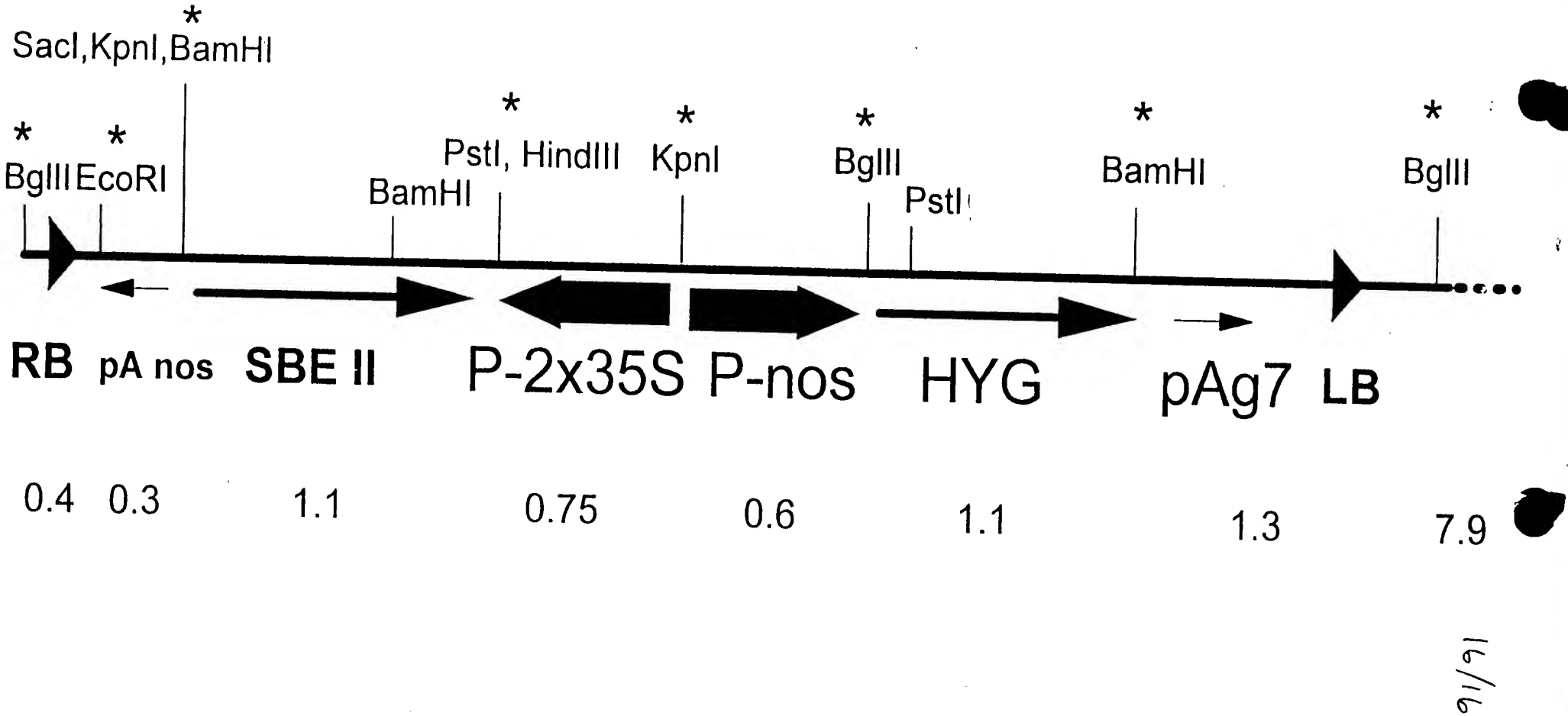

EcoR V
Bcl I
1100
1200
1300
1400
BamH I
Hinc II
1500
1600
Ssp I
Nsi I
Bcl I
1700
Sca I
Nco I
1800
Xmn I
1900
1919

Figure 11



9/16

① 97|030322

② 4 11 97

③ Keith Nash - 6